



# Interactions non-covalentes et propriétés physico-chimiques de petits systèmes biologiques : approches théoriques

Vanessa Riffet

## ► To cite this version:

Vanessa Riffet. Interactions non-covalentes et propriétés physico-chimiques de petits systèmes biologiques : approches théoriques. Chimie. Ecole polytechnique, 2014. Français. NNT : . tel-01084858

**HAL Id: tel-01084858**

**<https://pastel.archives-ouvertes.fr/tel-01084858>**

Submitted on 20 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

Présentée pour obtenir le grade de  
Docteur de l'École polytechnique

Spécialité : Chimie

Par

**Vanessa RIFFET**

# Interactions non-covalentes et propriétés physico-chimiques de petits systèmes biologiques : approches théoriques

Soutenue publiquement le 05 novembre 2014 devant le jury composé de :

Mme Élise Dumont	Maître de Conférences	Rapportrice
Mr Jean-Yves Salpin	Directeur de Recherche CNRS	Rapporteur
Mme Isabelle Demachy	Professeur des Universités	Présidente et Examinatrice
Mr Jean-Christophe Pouilly	Maître de Conférences	Examineur
Mr Gilles Frison	Chargé de Recherche CNRS	Directeur de thèse
Mr Guy Bouchoux	Professeur des Universités émérite	Co-directeur de thèse



## Remerciements

Au terme de ces trois années en tant que doctorante au sein du Laboratoire de Chimie Moléculaire (avant la fusion de 2014 le Laboratoire des Mécanismes Réactionnels (DCMR)), je souhaiterais remercier les personnes qui m'ont permis d'avancer.

Je tiens tout d'abord à remercier le professeur émérite Guy Bouchoux sans qui tout cela n'aurait pu être possible. Il y a quatre ans, je m'orientais vers une carrière d'enseignante en voulant passer l'agrégation de Chimie. Mais lors de mon stage de Master 2 en Chimie Organique avec le Professeur Guy Bouchoux, un vif intérêt pour la recherche a émergé. Ce fut une période d'éveil me concernant et j'ai décidé à la surprise générale de me lancer dans une thèse en chimie théorique co-encadrée par le Chargé de Recherche CNRS Gilles Frison, qui m'a appris à lancer mes premiers calculs et le Professeur émérite Guy Bouchoux.

Je remercie Gilles Ohanessian, directeur du DCMR de l'Ecole polytechnique de m'avoir accueillie dans son laboratoire de recherche.

La majorité de ce travail de thèse a principalement été encadré par Gilles Frison. Je te remercie de m'avoir guidé tout au long de cette thèse, de m'avoir permis de mieux appréhender l'utilisation de nombreux outils et méthodes de calculs nécessaires au bon déroulement de cette thèse et de la confiance que tu m'as accordé dans mon travail. J'ai également arrêté de stresser longtemps à l'avance pour des présentations orales, des dates limites et pour ce qui ne dépend plus de moi. Je me suis surprise à préparer un poster le matin même d'un congrès, ou encore à dire « il reste encore une heure, tout va bien, j'ai encore le temps ! ».

Un grand merci également à Guy Bouchoux pour la confiance qu'il m'a accordé lors de mon stage de Master 2. J'ai eu plaisir à travailler de nouveau avec vous sur la synthèse de Strecker dans cette thèse. Merci à vous de m'avoir écouté dans les moments difficiles et de m'avoir reboosté de temps de temps.

Je voudrais également remercier les collaborateurs dans ce travail. Tout d'abord Denis Jacquemin pour m'avoir accueilli dans son laboratoire le CEISAM, formé à la TD-DFT et



répondu à mes questions dans ce domaine ; mais aussi à Linda Fétéotava et Émilie Cauët pour les discussions scientifiques que j'ai pu avoir avec elles.

Merci à Carine Clavaguéra pour la gestion du cluster, sa disponibilité pour les soucis rencontrés avec les calculs et sa gentillesse.

Carine C., Éli R., Denis J., Florian T., Sophie B., Julia C.-G. et Tracy R., je vous dois un resto !!!

J'adresse mes plus sincères remerciements à Jean-Yves Salpin et Élise Dumont pour avoir accepté d'être les rapporteurs de mon travail de thèse. Merci également à Jean-Christophe Pouilly et Isabelle Demachy d'avoir accepté d'être les examinateurs de ce travail. Merci de l'intérêt que vous avez porté à mon travail de Recherche. J'espère que vous avez apprécié la lecture de ce manuscrit de thèse.

Merci à toutes les personnes que j'ai pu côtoyer pendant ces trois ans. Les permanents : Édith, Christophe, Sophie, Stéphane, Michel, Thérèse ; les post-doctorants : Ashwani, Manjeet, Éli ; les thésards : Julien, Aziz, Joe, Ahmad, Ophélie, Yannick, Florian, Jana et Madan. Les stagiaires : Pierre-Henri, Tatiana, Manali, ... Manali, j'ai apprécié encadrer ton stage lors de ma deuxième année de thèse ; ta motivation et ton investissement dans ton travail m'ont fait plaisir. Merci également aux nombreuses personnes que j'ai pu rencontrer et avec lesquelles j'ai pu discuter lors de divers congrès ...

Merci également à l'ensemble des joueurs de l'équipe de foot lors des tournois organisés par Xdoc et à Édith pour ses nombreux gâteaux même si notre équipe perdait la majorité de ces matchs.

Je n'ai pas oublié de remercier les personnes hors laboratoire !!!

Mickaël : Notre rencontre s'est faite au tournoi de foot de l'Ecole polytechnique. Depuis tu partages ma vie et je pense que tu es la plus belle rencontre que j'ai pu faire lors de ma thèse. Merci pour ton soutien au quotidien, ton humour et de t'avoir comme chauffeur ; le fait de ne plus pouvoir dormir dans la voiture le matin va énormément me manquer.

Merci à ma famille de m'avoir supporté et soutenu durant ma thèse et bien avant... ainsi qu'à mes amies que j'ai certes pas beaucoup vu pendant ces trois années, mais on va se rattraper ☺



# Sommaire

<b>INTRODUCTION GENERALE</b>	<b>11</b>
<b>CHAPITRE I : INTRODUCTION</b>	<b>13</b>
I.1. Les interactions non-covalentes	15
I.1.1. Interactions électrostatiques	15
I.1.2. Cas particuliers	19
I.2. Pourquoi s'intéresser aux interactions non covalentes ?	20
I.2.1. Interactions non-covalentes et critères structuraux	20
I.2.1.1. Les interactions non covalentes dans les systèmes biologiques	20
I.2.1.2. Le repliement des biomolécules : structure secondaire	21
I.2.2. Interactions non-covalentes et formation de complexes	23
I.2.2.1. Biomolécules hydratées	23
I.2.2.2. Biomolécules solvatées par des ions	24
I.2.3. Interactions non-covalentes et capture électronique	26
I.2.3.1. Les radiations ionisantes	26
I.2.4. Interactions non-covalentes et réactivité : cas des interactions avec H <sub>2</sub> O	27
I.2.4.1. H <sub>2</sub> O, un relai pour les transferts électroniques	27
I.2.4.2. H <sub>2</sub> O, un relai pour les transferts de proton	28
I.3. Spectrométrie de masse	28
I.4. Apport de la chimie théorique	32
I.5. Bibliographie	33
<b>CHAPITRE II : METHODOLOGIES</b>	<b>37</b>
II.1. Les méthodes ab initio	39
II.1.1. La méthode Hartree-Fock	40
II.1.2. Les méthodes post Hartree-Fock	43
II.1.2.1. Interaction de configurations	43
II.1.2.2. Le champ multi-configurationnel auto-cohérent (MCSCF)	44
II.1.2.3. Complete Active Space Self-Consistent Field (CASSCF)	44
II.1.2.4. Interaction de configurations multi-référence (MRCI)	45
II.1.2.5. Coupled Cluster (CC)	46
II.1.2.6. La théorie des perturbations : Møller-Plesset	46
II.2. La théorie de la fonctionnelle de la densité	48
II.2.1. Théorie de la Fonctionnelle de la Densité indépendante du temps (DFT)	48
II.2.1.1. Généralités	48
II.2.1.2. L'approximation de la densité locale (LDA)	50
II.2.1.3. L'approximation du gradient généralisé (GGA)	51
II.2.1.4. Les fonctionnelles meta-GGA	52
II.2.1.5. Les fonctionnelles hybrides	53
II.2.1.6. Les fonctionnelles corrigés à longue portée	53
II.2.1.7. Les effets de dispersion	54
II.2.2. Théorie de la Fonctionnelle de la Densité Dépendante du temps (TD-DFT)	56
II.3. Les bases d'orbitales atomiques	58
II.4. Les méthodes composites Gaussian-n	59
II.5. Les méthodes semi-empiriques	61

II.5.1. Généralités	61
II.5.2. Méthodes PM6 et PM6-DH2	62
II.6. Exploration des surfaces d'énergie potentielle	63
II.6.1. Exploration en mécanique classique	63
II.6.1.1. Les outils	63
II.6.1.1.1. Généralités	63
II.6.1.1.2. AMOEBA, Atomic Multipole Optimized Energetics for Biomolecular Applications.	65
II.6.1.1.3. Dynamique moléculaire classique	68
II.6.1.1.4. Simulation Monte Carlo	69
II.6.1.1.5. Saut de bassins	69
II.6.1.2. Stratégies d'exploration de surface pour les systèmes micro-hydratés	70
II.6.1.2.1. L'approche arbre de Darwin	70
II.6.1.2.2. L'approche hiérarchique	71
II.6.2. Les Chemins Réactionnels	72
II.6.3. Remarques	72
II.7. Les méthodes hybrides QM/MM	73
II.7.1. Généralités	73
II.7.2. L'approche ONIOM	73
II.8. Les effets de solvation	74
II.9. Outils d'analyse	75
II.9.1. Les grandeurs thermodynamiques	75
II.9.2. La méthode NCI	78
II.9.3. Etude de la densité de spin	79
II.9.3.1. Densité de spin SCF	79
II.9.3.2. Densité de spin atomique	79
II.9.4. L'analyse de décomposition en énergie	80
II.9.5. Erreur de superposition de base	80
II.10. Bibliographie	82

## **CHAPITRE III : EFFETS DES INTERACTIONS NON-COVALENTES SUR LES PROPRIETES STRUCTURALES ET ENERGETIQUES** **87**

### **Partie I : Influences réciproques des interactions non-covalentes des acides aminés acétylés micro-hydratés** **89**

III.1.1. Introduction	91
III.1.2. Exploration de la surface d'énergie potentielle	96
III.1.2.1. Couplage des approches hiérarchique et arbre de Darwin	96
III.1.2.2. Amélioration de l'approche hiérarchique	111
III.1.3. Influences réciproques des interactions non-covalentes	122
III.1.4. Conclusions	142
III.1.5. Bibliographie	144

### **Partie II : Clusters de glycine bêtaïne en phase gazeuse** **145**

III.2.1. Introduction	147
III.2.1.1. Définition de la glycine bêtaïne	147
III.2.1.2. Travaux antérieurs	147
III.2.1.2.1. Exemple de Clusters étudiés dans la littérature	147
III.2.1.2.2 Clusters de glycine bêtaïne dans la littérature	149
III.2.2. Stratégie d'étude	152

III.2.2.1. Stratégie d'exploration de la surface d'énergie potentielle	152
III.2.2.2. Base spécifique pour $[\text{GB}_{13} + 2\text{H}]^{2+}$	157
III.2.3. Résultats et discussion	158
III.2.3.1. Structures géométriques	158
III.2.3.2. Nature des interactions non-covalentes intermoléculaires	165
III.2.3.2.1. Interactions non-covalentes dans $[\text{GB}_n + 2\text{H}]^{2+}$	165
III.2.3.2.2. Cohésion au sein des clusters $[\text{GB}_n + 2\text{H}]^{2+}$	169
III.2.3.3. Processus de fragmentations	173
III.2.3.3.1. Évaporation d'une glycine bêtaïne neutre	173
III.2.3.3.2. Explosion coulombienne des clusters $[\text{GB}_n + 2\text{H}]^{2+}$	175
III.2.3.3.3 Bilan	178
III.2.4. Conclusions et Perspectives	179
III.2.5. Bibliographie	181
<b>CHAPITRE IV : EFFETS DE L'AUTO-SOLVATATION SUR LES PROPRIETES ELECTRONIQUES DE PEPTIDES RADICAUX CATIONS</b>	<b>183</b>
IV.1. Introduction	185
IV. 2. Stratégie d'analyse des structures électroniques	188
IV.3. Choix des fonctionnelles en DFT et TD-DFT pour l'étude des états fondamental et excités de peptides radicaux cations riches en hydrogène.	190
IV.4. Les liaisons hydrogène, un facteur clé de la structure électronique des peptides radicaux cations	208
IV.4.1. Etudes de systèmes modèles.	209
IV.4.2 Applications	211
IV.5. Conclusions	214
IV.6. Bibliographie	215
<b>CHAPITRE V. EFFETS DES LIAISONS HYDROGENE SUR DES PROCESSUS DE TRANSFERT DE <math>\text{H}^+</math> ET/OU DE <math>\text{H}^\bullet</math></b>	<b>217</b>
<b>Partie I. Formation de l'intermédiaire radical cation</b>	<b>219</b>
V.1.1. Généralités	221
V.1.2. Etude de la relaxation de peptides radicaux cations	222
V.1.3. Conclusions	229
V.1.4. Bibliographie	231
<b>Partie II. Effet de la micro-solvatation sur les premières étapes de la réaction de Strecker. Exemple de la glycine</b>	<b>233</b>
V.2.1. La chimie prébiotique	235
V.2.3. Stratégies	239
V.2.4. Résultats et discussion	241
V.2.4.1. Les premières étapes de la synthèse de Strecker dans un milieu neutre	242
V.2.4.3. Les premières étapes de la synthèse de Strecker dans un milieu acide	258
V.2.5. Conclusion	270
V.2.6. Bibliographie	271
<b>CONCLUSIONS GENERALES ET PERSPECTIVES</b>	<b>273</b>



# Introduction générale

La structure primaire des peptides et protéines est régie par les liaisons covalentes entre les divers résidus d'acides aminés tandis que leur structuration spatiale est gouvernée par diverses interactions non-covalentes, inter- et intra-moléculaires. Les propriétés biologiques (catalyse enzymatique, transport moléculaire...) et physico-chimiques (solubilité, acido-basicité, oxydo-réduction, réponse aux radiations ionisantes...) de ces systèmes biologiques dépendent à la fois des structures primaires et secondaires, et par conséquent des enchainements de liaisons de covalence et des interactions de type liaisons hydrogène et hydrophobes. Actuellement, de nombreuses études sont menées afin de décrire, de mesurer et de comprendre les effets des interactions non-covalentes sur ces systèmes. Dans ce cadre, ce travail de thèse vise à une meilleure compréhension de l'effet de ces interactions au sein de petits peptides et d'agrégats. Ces effets ont été classés en trois catégories que nous présentons dans la suite successivement : (i) l'effet sur la structure tridimensionnelle d'une molécule ou d'un agrégat (chapitre III) ; (ii) l'effet sur la structure électronique d'une molécule (chapitre IV) ; (iii) l'effet sur la réactivité et la fragmentation de molécules (chapitre V). Les systèmes chimiques d'intérêt biologique que nous avons étudié sont multiples, permettant ainsi, au cours de nos travaux, de couvrir une grande diversité de situations et des effets des interactions non-covalentes.

Dans un premier chapitre, nous présentons de manière succincte ce que recouvre la dénomination "interactions non-covalentes", leurs intérêts et les notions de spectrométrie de masse utile à la compréhension de nos travaux.

Toutes nos études reposent sur l'utilisation des outils de la chimie quantique et de la mécanique moléculaire. Les principales méthodes utilisées dans ce travail de thèse sont décrites dans le Chapitre II. Du fait de la complexité et de la taille des systèmes biologiques réels, nous avons élaboré et étudié des modèles chimiques simplifiés. Une approche partant de la phase gazeuse pour aboutir à des espèces microhydratées, nous a permis d'étudier la nature et les effets d'interactions non-covalentes prises « individuellement » sur des petits peptides. Une étude similaire a été réalisée sur des agrégats de glycine-bétaïne. Le champ de force polarisable AMOEBA a été utilisé pour produire un premier ensemble de structures d'intérêts. Les méthodes semi-empiriques, la Théorie de la Fonctionnelle de la Densité (DFT), les méthodes composites et post-Hartree-Fock nous ont permis d'affiner les recherches



conformationnelles, les géométries et/ou les calculs thermochimiques. La DFT dépendante du temps (TD-DFT) a été indispensable pour étudier les états excités de systèmes radicalaires après capture électronique.

Dans le troisième chapitre, nous présentons des résultats concernant l'effet des interactions non-covalentes sur l'agrégation de systèmes chargés positivement. Les groupes chargés ammonium, guanidinium ou imidazolium, sont présents dans les chaînes latérales des acides aminés lysine, arginine et histidine protonés. La solvation par un nombre discret de molécules d'eau a été examinée sur les modèles peptidiques constitués des acides aminés Lysine, Arginine et Histidine N-acétylés et protonés. Le processus d'agrégation a été étudié sur une molécule porteuse d'un ammonium quaternaire: la glycine bétaine. Pour ces deux types d'études, il a été nécessaire de mettre au point des protocoles d'exploration de surfaces d'énergie potentielle à partir du champ de force polarisable AMOEBA, ceci afin d'optimiser l'exhaustivité des recherches conformationnelles pour ces systèmes biologiques flexibles.

Dans le chapitre IV, nous avons tout d'abord réalisé une calibration des méthodes de modélisation adaptées à l'étude des propriétés électroniques d'espèces radicalaires. Nous avons ensuite abordé l'effet structurant des liaisons hydrogène sur des peptides di-chargés subissant une réduction électronique.

Le chapitre V a un contexte différent des autres études car il concerne l'incidence des interactions non-covalentes sur la réactivité. Tout d'abord, nous avons examiné le rôle de la structure électronique et des interactions non-covalentes lors des processus de transfert et d'élimination d'atome d'hydrogène qui interviennent lors des processus de dissociation par capture ou transfert d'électron (ECD et ETD). Ensuite, l'effet d'un nombre croissant de molécules d'eau sur les barrières d'activation des deux premières étapes de la synthèse de Strecker a été entrepris. Comme pour le chapitre III, une large exploration de la surface d'énergie potentielle de chaque système a été réalisée. De plus, une proposition de scénario a été faite quant à un mécanisme possible pour la formation des acides aminés dans le milieu prébiotique.

# Chapitre I : Introduction



Les liaisons covalentes résultent de la mise en commun des électrons de la couche électronique la plus externe de deux atomes. La liaison formée est forte ce qui implique des faibles longueurs de liaisons et des énergies de liaisons de l'ordre de la centaine de  $\text{kJ mol}^{-1}$ . Si la différence d'électronégativité entre les deux atomes d'une liaison est importante, on parlera alors de liaison polarisée ou à caractère ionique partiel (HCl par exemple).

Les autres types d'interactions liantes ou non liantes entre atomes sont classés dans la catégorie des interactions non-covalentes ; il s'agit principalement d'interactions électrostatiques que nous allons présenter brièvement. Ces interactions, inter et intra-moléculaires, sont d'autant plus nombreuses que les molécules comportent un grand nombre d'atomes et (ou) sont flexibles. Elles sont essentielles pour la structuration des peptides et des clusters, et interviennent également dans la plupart des processus réactionnels en phase gaz ou en solution comme les transferts de proton, d'électron et la capture électronique.

## I.1. Les interactions non-covalentes

Cette section a été inspirée de trois ouvrages.[1-3]

### *I.1.1. Interactions électrostatiques*

Il existe plusieurs types d'interactions électrostatiques impliquant des ions et des dipôles. Parmi elles, les interactions de type van der Waals sont des interactions électrostatiques en  $\frac{1}{r^6}$  entre dipôles. Elles englobent trois catégories d'interactions : les interactions de type Keesom, de Debye et de London.

#### **Interactions ioniques**

L'interaction électrostatique la plus connue est celle qui existe entre deux espèces chargées positivement ou négativement. Cette interaction peut être répulsive si les charges sont de même signe ou attractive si elles sont de signes opposés. Charles-Augustin Coulomb a proposé au début du XVIII<sup>ème</sup> siècle une loi permettant d'estimer le gain (-) ou la perte (+) d'énergie lorsque l'on rapproche deux charges ponctuelles  $z_1e$  et  $z_2e$  de l' $\infty$  à la distance  $r$  :

$$E = \pm \frac{z_1 z_2 e^2}{(4\pi \epsilon_0) \epsilon r} \quad (\text{I.1})$$

où  $\epsilon_0$  est la permittivité du vide et  $\epsilon$  est la constante diélectrique du milieu. Lorsque les deux charges sont portées par des atomes (ou des ions), on parle de liaisons ioniques en phase solide.

### Interactions ion – dipôle

Une charge électrique ponctuelle et un dipôle permanent subiront une interaction électrostatique dont l'énergie est fonction de la distance  $r$  entre la charge et le centre de gravité du dipôle et de l'angle  $\theta$ .

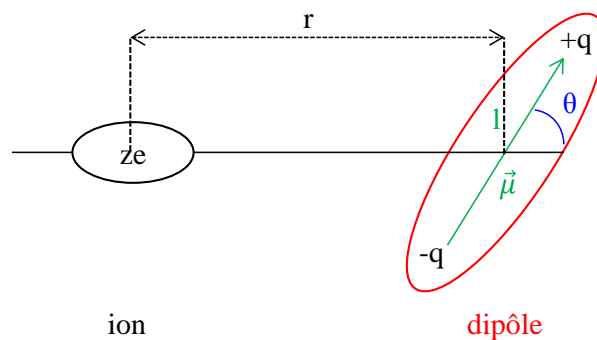


Figure I.1. Caractéristiques pour une interaction de type ion – dipôle ( $l \ll r$ ).

Cette énergie d'interaction a pour expression :

$$E = \pm \frac{ze\mu \cos \theta}{4\pi\epsilon_0\epsilon r^2} \quad (\text{I.2})$$

et peut prendre des valeurs positives ou négatives selon l'orientation du dipôle. Elle dépend du moment dipolaire  $\mu$  qui est la grandeur physique égale au produit de la valeur absolue des charges  $q$  par la distance qui les séparent ( $\vec{\mu} = q\vec{l}$ ). Toutefois, lorsque l'énergie d'interaction, définie à l'équation (I.2), est similaire à l'énergie thermique  $kT$ , le dipôle est dans un état de rotation thermique naturelle. Par conséquent, l'énergie coulombienne comprend seulement la probabilité d'orientation par rapport à la charge ponctuelle. Par exemple, ce type d'interaction joue un rôle important lors de la solvation des ions par un solvant polaire (eau, éthanol, ammoniac, acétonitrile,...).

### Interactions ion – dipôle induit

Un champ électrique  $E_{\text{elec}}$  créée par un ion va perturber la distribution de charges au sein d'un système apolaire. Cette déformation du nuage électronique du système apolaire est mesurée par une grandeur physique que l'on nomme polarisabilité et est notée  $\alpha$ . Le dipôle induit  $\mu_{\text{induit}} = \alpha E_{\text{elec}}$  sera orienté dans le sens opposé du champ électrique et interagira avec lui de manière attractive. Cette interaction est toujours favorable et sera d'autant plus forte que la polarisabilité du système apolaire est grande. L'énergie d'interaction associée a comme expression :

$$E = -\frac{(ze)^2\alpha}{2(4\pi\epsilon_0\epsilon)^2r^4} \quad (\text{I.3})$$

### Interactions dipôle – dipôle

Lorsque deux dipôles fixes interagissent (Figure I.2), l'énergie de leur interaction s'obtient en considérant toutes les contributions coulombiennes entre les charges ponctuelles.

$$E = \pm \frac{\mu_1\mu_2}{4\pi\epsilon_0\epsilon r^3} [2 \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \cos \varphi] \quad (\text{I.4})$$

Cette énergie est fortement dépendante de l'orientation des deux dipôles qui est définie par les angles  $\theta_1$ ,  $\theta_2$  et  $\varphi$ .

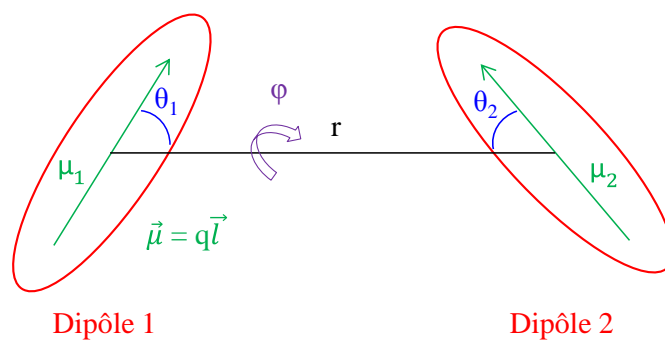


Figure I.2. Caractéristiques pour une interaction de type dipôle – dipôle.

Or les dipôles ne sont pas fixes en raison d'une rotation thermique naturelle de ces dipôles. Dans ce cas, l'interaction résultante se nomme interaction de Keesom.

La liaison de type  $AX \cdots BR$  est un cas particulier de l'interaction de type Keesom ou de l'interaction ion – dipôle. La Figure I.3 présente les caractéristiques de ce type d'interaction

essentiellement localisée entre X et B i.e. les deux charges de signes opposés des deux dipôles  $\delta^+$  et  $\delta^-$ . Dans le cas où  $X = H$ , on parlera alors de liaison hydrogène, lorsque  $X = \text{halogène}$  il s'agira de liaison halogène. La charge partielle positive,  $\delta^+$ , est portée par l'atome d'hydrogène ou d'halogène. La charge ou le second dipôle serait B qui possède un ou plusieurs doublets non liants et une charge partielle  $\delta^-$ . A et B sont donneur et accepteur de liaison hydrogène (ou halogène) respectivement. La force de la liaison hydrogène ou halogène va dépendre de (i) la distance entre X et B, (ii) de l'angle  $A-X\cdots B$  et (iii) des électronégativités des atomes.

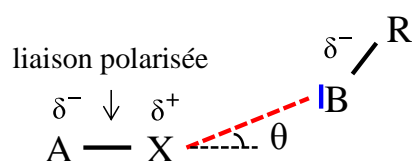


Figure I.3. Caractéristiques pour une interaction de type  $AX\cdots BR$ , tel que  $X = H$  ou halogène.

### Interactions dipôle – dipôle induit

Un raisonnement similaire à l'interaction ion – dipôle induit peut être fait dans le cas de l'interaction dipôle – dipôle induit mais qui est moins forte que l'interaction ion – dipôle induit. En considérant la rotation thermique naturelle du dipôle, l'interaction résultante se nomme interaction de Debye et l'énergie associée a cette expression :

$$E = -\frac{\mu^2 \alpha (1 + 3 \cos^2 \theta)}{2(4\pi\epsilon_0\epsilon)^2 r^6} \quad (I.5)$$

Ces interactions sont toujours favorables et seront d'autant plus fortes que le moment dipolaire du dipôle et la polarisabilité du système apolaire sont grands.

### Interactions dipôle instantanée – dipôle induit

Ce type d'interaction se nomme interaction de dispersion de London. Dans une molécule, lorsque la distribution moyenne des moments dipolaires est nulle, la distribution de charge n'est pas nécessairement statique. En effet, à un instant  $t$ , il existe des moments dipolaires instantanés résultant de l'inhomogénéité de la distribution électronique dans une molécule. Ce dipôle induit instantanément crée un champ électrique qui va polariser les atomes neutres

voisins et (ou) la molécule neutre voisine. Plus la polarisabilité des systèmes apolaires est grande, plus l'interaction de London sera forte.

### ***1.1.2. Cas particuliers***

#### **Interactions de recouvrement**

Lorsque deux atomes (molécules) s'approchent l'un(e) de l'autre, il y a interpénétration de leurs nuages électroniques. D'après le principe d'exclusion de Pauli, des spins de même sens ne peuvent pas occuper la même position dans l'espace. Ainsi, ils se repoussent mutuellement. Ce phénomène correspond à la répulsion de Pauli. Lorsque la distance entre les deux systèmes devient très faible, alors la répulsion de Pauli prédomine sur l'interaction électrostatique.

#### **Interaction hydrophobe**

Dans un milieu aqueux, les molécules/groupements hydrophobes se regroupent pour minimiser le contact avec les molécules d'eau. Les molécules de H<sub>2</sub>O vont alors former une cavité autour des molécules hydrophobes afin de créer un nombre maximum de liaisons hydrogène entre elles, ce qui entraîne localement une augmentation de la structuration de l'eau. Ce processus se nomme hydratation hydrophobe. Puisque le système résultant est plus ordonné en présence des molécules de H<sub>2</sub>O, ce processus est enthalpiquement favorable mais entropiquement défavorable. Parce que les molécules/groupements hydrophobes sont enveloppées par les molécules de H<sub>2</sub>O, des attractions favorables (van der Waals,  $\pi \cdots \pi$ , ... ) entre les groupements apolaires vont s'exercer.



## I.2. Pourquoi s'intéresser aux interactions non covalentes ?

### I.2.1. Interactions non-covalentes et critères structuraux

#### I.2.1.1. Les interactions non covalentes dans les systèmes biologiques

Il est admis que la structure tridimensionnelle, les propriétés chimiques et indirectement le rôle des polymères biologiques tels que les peptides sont gouvernés non seulement par leur composition primaire mais aussi par diverses interactions non-covalentes. En fonction du pH physiologique, la protéine peut exister sous différentes formes comme multichargée en général par  $\text{nH}^+$  + zwitterionique par exemple. La Figure I.4 présente la « forme zwitterionique » d'un peptide.



Figure I.4. Forme zwitterionique d'un peptide.

Pour ces formes citées, certains groupements chimiques situés dans les chaînes latérales et les  $\text{N}_{\text{term}}$  et  $\text{O}=\text{C}_{\text{term}}$  sont susceptibles d'être protonés ou déprotonés. Ces groupements deviennent alors chargés et permettent de créer des interactions de type ion – ion, ion – dipôle et certaines liaisons hydrogène, facteurs déterminant d'un point de vue structural de la protéine. En plus des différents sites chargés, les protéines présentent un grand nombre de dipôles, comme celle représentée dans la Figure I.5,

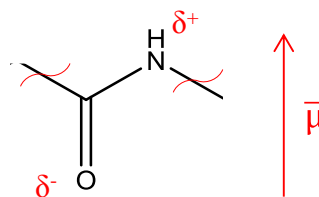


Figure I.5. Liaison peptidique

ou ceux présents sur les chaînes latérales des fonctions OH, phénol, SH, cycle indole, COOH, ... De là, les interactions de types dipôle – ion, dipôle – dipôle et dipôle – dipôle induit sont possibles. Les interactions de type dipôle instantané – dipôle induit se font entre les groupements apolaires des chaînes latérales, comme les groupes alkyles voire les cycles aromatiques pour des interactions de types  $\pi$ -stacking des protéines. Les interactions de type cation $\cdots\pi$  (cation = LysH<sup>+</sup> et ArgH<sup>+</sup> et  $\pi$  = Phe, Tyr, ou Trp par exemple) sont très souvent rencontrées dans les protéines. Dougherty et son équipe ont montré que (i) toutes les protéines de grandes tailles possèdent au moins une interaction de type cation $\cdots\pi$ , (ii) un quart des groupements indole provenant du Trp interviennent dans ce type d'interaction avec ArgH<sup>+</sup> ou LysH<sup>+</sup> et (iii) que le groupement guanidinium participe davantage à ce type d'interaction que l'ammonium de la lysine.[4] D'autres interactions de type  $\pi$  sont possibles dans la zone hydrophobe des biomolécules, comme les interactions de type  $\pi\cdots\text{HC}$ .

La connaissance des sites de protonation des acides aminés est importante pour comprendre certains processus biologiques, comme la photosynthèse[5] ou la respiration cellulaire[6] puisqu'ils impliquent des transferts de protons au sein des protéines. La détermination des acidités et basicités intrinsèques des acides aminés isolés et des peptides est essentielle pour localiser les différents sites de protonation et comprendre l'ensemble de ces processus biologiques. C'est pourquoi, que ce soit expérimentalement ou théoriquement, la mesure et le calcul de ces données intrinsèques intéressent un grand nombre de chercheurs.[7-14]

Les composés halogénés sont utilisés dans le domaine thérapeutique. Par le biais de la cristallographie par rayons X (partie 1.3.1) des complexes de type protéine $\cdots$ ligand halogéné (=récepteur $\cdots$ médicament), il a été prouvé que ce type d'interaction est bien présent dans les systèmes biologiques,[15,16] et qu'il est important pour la spécificité de la reconnaissance moléculaire[15,17].

### I.2.1.2. Le repliement des biomolécules : structure secondaire

Dans les peptides, les chaînes (poly)peptidiques se structurent grâce aux différentes interactions non-covalentes inter- et intramoléculaires citées précédemment. La structure secondaire résultante peut avoir la forme d'une hélice, d'un feuillet voire d'un « globule » (structure approximativement sphérique). La localisation et le nombre de liaisons hydrogène existantes entre les groupements carbonyles et les NH amide permettent de stabiliser ces

diverses structurations, ce qui fait de cette interaction  $\text{-NH}\cdots\text{O=}$  la plus importante dans les systèmes biologiques. De nombreuses études théoriques et expérimentales ont pour objectifs de comprendre la structuration en hélice en fonction de la nature des interactions, du nombre de résidu, de l'état de charge et/ou de la nature de la charge (cation métallique ou proton).[18-28] Les modèles utilisés sont souvent des polyacides aminés qui adoptent soit une structure en hélice, soit une structure globulaire. Par exemple, une étude conjointe entre expérience et théorie de polyalanines complexées par  $\text{Na}^+$  a montré que ces deux types de structures coexistent, ce qui est le cas pour l'octa-peptide  $\text{Ala}_n\text{Na}^+$  avec  $n = 8$  où la structure en hélice (b) est plus stable de seulement  $2.6 \text{ kJ mol}^{-1}$  par rapport à la forme globulaire (a) (voir Figure I.6).[26]

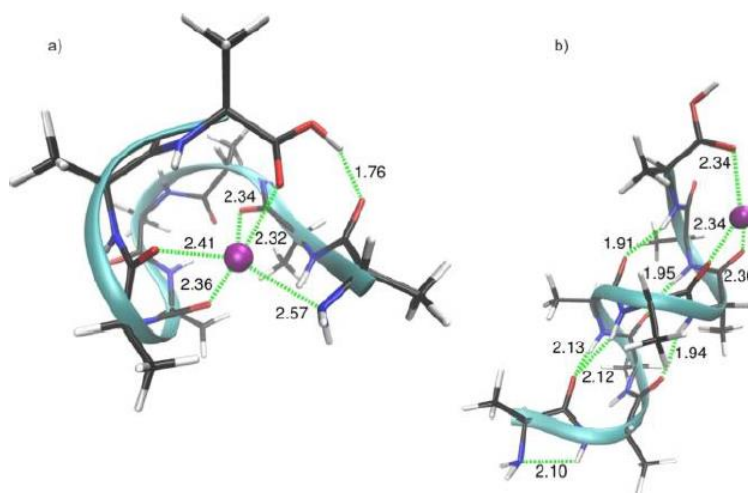


Figure I.6.  $\text{Ala}_8\text{Na}^+$  : (a) structure globule la plus basse en énergie ; (b) structure  $\alpha$ -hélice la plus basse en énergie. (M06/6-311+G(d,p)//M06/6-31G(d,p). Cette image est extraite de la référence [26].

À partir de  $n = 9$ , la forme hélice est majoritaire. Kohtani et al. ont montré grâce à des mesures de mobilité ionique en fonction de la température qu'il existe une interconversion rapide entre les structures globule et hélice au-delà de  $T = 450 \text{ K}$ . Cette interconversion est accompagnée d'un transfert de proton le long de l'hélice.[29] De plus, l'agencement des acides aminés au sein du polyacide aminé est également important pour la structuration. En effet, les peptides protonés  $\text{Ac-K(AGG)}_5\text{H}^+$  et  $\text{Ac-LysH}^+\text{-Ala}_n$  adoptent une structure globulaire pour  $n < 13$ , tandis que  $\text{Ac-(AGG)}_5\text{K}^+\text{H}^+$  et  $\text{Ac-Ala}_n\text{-LysH}^+$  pour  $n \geq 7$  ont une structure en hélice.[19,22]

## ***1.2.2. Interactions non-covalentes et formation de complexes***

### **1.2.2.1. Biomolécules hydratées**

Le repliement des protéines est dépendant à la fois des interactions intramoléculaires et des interactions intermoléculaires entre la séquence d'acides aminés et leur environnement. Dans leur milieu biologique habituel, les protéines sont dans un environnement aqueux. Les interactions entre des molécules polaires et des groupements hydrophobes sont défavorables, ce qui implique le regroupement des groupements hydrophobes maintenu par des interactions de type dipôle instantané – dipôle induit au milieu de la protéine minimisant ainsi leur contact avec les molécules de H<sub>2</sub>O. Les groupements polaires se positionnent en direction de l'eau et forment des liaisons hydrogène avec cette dernière. L'eau est considérée comme une molécule participant dans certains processus biologiques et dans la structuration des peptides et n'est plus vue comme un environnement inerte.[30]

Afin de comprendre les différences physico-chimiques existant entre les peptides solvatés et isolés, la solvation de ces systèmes par un nombre donné de molécules d'eau est encore aujourd'hui beaucoup étudiée. L'étude de la micro-hydratation des molécules isolées permet de comprendre le rôle des premières molécules de H<sub>2</sub>O sur les propriétés physico-chimiques (thermochimiques et structurales) des acides aminés et des peptides.[31,32] Plusieurs processus sont analysés comme par exemple : (i) le transfert de proton permettant de passer de la forme neutre à la structure zwitterionique (transition phase gaz – phase aqueuse) ;[33-38] (ii) la durée de vie des états excités ;[39] (iii) la modification de la structure secondaire comme observé par exemple pour le N-acetyl-phenylalanine-amide ;[40] (iv) la propension d'une structure hélice ;[41] ou encore (v) la fixation sélective de molécule de H<sub>2</sub>O sur des systèmes de même composition mais de structure secondaire différente.[40,42-44] De plus, certaines molécules de H<sub>2</sub>O de la première sphère d'hydratation participent de façon conséquente à la structuration des peptides puisqu'elles cristallisent avec des molécules de H<sub>2</sub>O, comme le montre par exemple la structure établie par spectroscopie des rayons X de la protéine villin headpiece subdomain représentée dans la Figure I.7.[45,46]

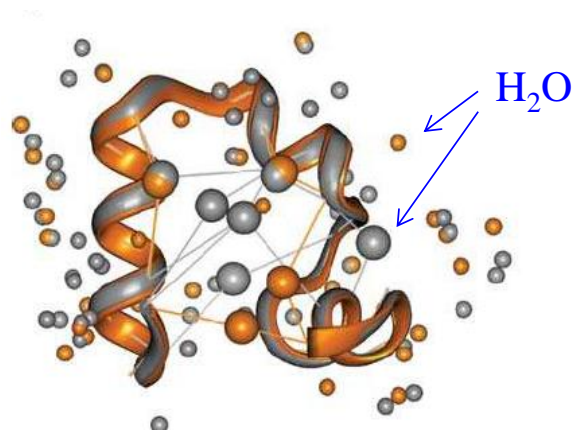


Figure I.7. Structures Rayon-X à haute résolution de la villin headpiece subdomain à pH 5.1 (orange, 1.55 Å de résolution, code PDB 1WY4) et pH 7.0 (gris, 0.95 Å de résolution, code PDB 1WY3). Cette image est extraite de la référence [45].

## I.2.2.2. Biomolécules solvatées par des ions

### **solvatation par des (alkyl)ammoniums**

Au vu de la composition variée de la structure primaire des protéines, il y a d'autres facteurs que l'eau qui influencent la stabilisation de la structure zwitterionique. T. B. McMahon et ses collaborateurs ont étudié les interactions entre différents acides aminés neutres et des ions ammonium[47-49] ou d'alkylammonium[48,50] se produisant en phase gazeuse. Cette interaction permet de mimer par exemple l'interaction entre un C=O et un ammonium provenant d'une lysine. Au cours de leurs travaux, ils ont déduit de spectres de dissociation multiphotonique infra-rouge (IRMPD) que la proline et la glycine zwitterionique sont bien stabilisés par un méthylammonium et un ammonium respectivement. De plus, ils ont montré que lorsque l'affinité protonique des acides aminés aliphatiques augmente, (Gly, Ala et Val), la structure zwitterionique est davantage stabilisée ; et inversement, lorsque l'affinité protonique des (alkyl)ammoniums augmente, l'interaction avec un acide aminé sous forme zwitterionique est plus faible.[48]

### **Solvatation par un cation métallique**

La présence d'un cation peut avoir un rôle déterminant dans (i) les propriétés structurales,[51,52] (ii) les transferts d'électron, (iii) la fixation et le transport de O<sub>2</sub>, (iv) la fixation, l'activation et le renouvellement du substrat et (v) le stockage et le relargage des métaux.[53] Afin de pouvoir comprendre ces effets, de nombreuses études théoriques et expérimentales portent sur la complexation entre les acides aminés (et leur dérivés) et différents cations métalliques en phase gazeuse. [54-70] Il a été mis en évidence que le(s)

site(s) de chélation du cation métallique sur les acides aminés dépendent de plusieurs paramètres : la nature du cation métallique, la nature de la chaîne latérale (aromatique, aliphatique, oxygénée, ...) et le nombre de sites disponibles. En fonction du cation, l'acide aminé peut adopter la forme zwitterionique comme forme la plus stable.[59-65,68,70]

Des études sur la complexation des acides aminés (déprotonés ou non) par des cations métalliques toxiques tels qu' $\text{Al}^{3+}$ [71] et  $\text{Pb}^{2+}$ [72-75] ont été faites afin de mieux comprendre les interactions avec les protéines, d'identifier les sites de fixation du cation métallique et le rôle de l'hydratation sur ces sites (voir partie « Solvation par un cation métallique et  $n\text{H}_2\text{O}$  »). Ces études sont motivées par le fait que certains cations métalliques sont associés à la formation d'agrégats toxiques qui peuvent être impliqués dans les maladies neurodégénératives.[76,77] Par exemple, Burt et ses collaborateurs ont montré que pour divers acides aminés déprotonés (Ala, Val, Leu, Ile et Pro),  $\text{Pb}^{2+}$  interagit avec  $\text{HN}^-$  et l'oxygène du carbonyle ; pour la lysine, déprotoné sur le groupement acide carboxylique, l'amine de la chaîne latérale participe à la stabilisation de ce cation.

Outre les cations métalliques de type  $\text{M}^{z/+}$ , les organostanniques peuvent également influencer la stabilité relative des deux formes des acides aminés.[78]

### **Solvation par un cation métallique et $n\text{H}_2\text{O}$ .**

Les métalloprotéines sont actives dans un environnement aqueux, il est donc important d'examiner les effets combinés du cation métallique, des chaînes latérales de la protéine et de quelques molécules de  $\text{H}_2\text{O}$  sur la structure et l'activité biologique de ces biomolécules. Les systèmes étudiés sont en général mimés par des petits modèles tels que les acides aminés et des dérivés de petites tailles. De nombreuses études ont examiné le rôle du cation métallique et (ou) du nombre de molécules de  $\text{H}_2\text{O}$  sur les formes canoniques et zwitterioniques.[59-62,65-67,79-84] Par exemple, la micro-hydratation a un effet sur la stabilité relative des complexes de type  $\text{AA}\cdots\text{Ca}^{2+}$  par rapport à ces mêmes complexes isolés, puisque pour les complexes non-hydratés l'ordre est le suivant :  $\text{Phe}\cdots\text{Ca}^{2+} > \text{Thr}\cdots\text{Ca}^{2+} > \text{Gly}\cdots\text{Ca}^{2+}$ , et une fois hydratée avec 6 à 8  $\text{H}_2\text{O}$ ,  $\text{Thr}\cdots\text{Ca}^{2+}(\text{H}_2\text{O})_n$  a la meilleure stabilité :  $\text{Thr}\cdots\text{Ca}^{2+}(\text{H}_2\text{O})_n > \text{Phe}\cdots\text{Ca}^{2+}(\text{H}_2\text{O})_n > \text{Gly}\cdots\text{Ca}^{2+}(\text{H}_2\text{O})_n$ .[85]

### **Solvation par un anion**

Les interactions de type cation...peptides ont été intensivement étudiées. Celles avec les anions sont bien moins connues et peu d'études théoriques et expérimentales sont à ce jour présentes dans la littérature. En 2000, Simons et son équipe ont montré que la glycine

zwitterionique réduite est stabilisée de  $\sim 38 \text{ kJ mol}^{-1}$  par rapport à la forme canonique réduite, ce qui est inférieur à l'écart entre ces deux formes non réduites.[86] Depuis, les équipes de Williams et Dessent se sont penchées sur la formation de la forme zwitterionique de certains acides aminés en présence d'ions halogénures. Dessent et son équipe ont réalisé une étude de Dissociation Induite par Collision (CID) sur des systèmes de type  $X^{\cdots}\text{Arg}$  et ont montré que pour  $X^- = \text{Br}^-$  et  $\text{NO}_3^-$ , ces systèmes sont sous forme zwitterionique.[87] Des spectres IRMPD sur Glu, His et Arg et IRPD pour Pro complexés par un ion halogénure ont également été enregistrés.[88,89]  $X^{\cdots}\text{Glu}$  et  $X^{\cdots}\text{His}$  (avec  $X=\text{Cl}$  et  $\text{Br}$  et  $\text{I}$ ) existent uniquement sous leur forme canonique tandis que  $X^{\cdots}\text{Arg}$  ( $X=\text{Cl}$  et  $\text{I}$ ) est sous forme zwitterionique. Des calculs effectués au niveau de théorie MP2/6-311+G(d,p) montrent que les formes zwitterionique et canonique coexistent pour  $\text{Br}^{\cdots}\text{Arg}$  mais que seule la forme canonique est majoritaire pour  $\text{Cl}^{\cdots}\text{Arg}$ .[90] Ce résultat est en contradiction avec l'étude menée par l'équipe de Williams. Mais les travaux sur  $\text{Cl}^{\cdots}\text{Pro}$  de Kass soulignent l'importance de l'aspect cinétique lors de la fragmentation.[91] En effet, les formes zwitterionique et canonique sont iso-énergétiques mais seule la forme canonique est détectée dans les spectres IRPD car la forme zwitterionique se fragmente plus lentement. D'autres études théoriques portant sur la stabilisation de la glycine zwitterionique par des mono- et di-anions (oxalate et malonate) se trouvent dans la littérature.[92,93] Parmi les structures proposées les anions se coordonnent sur les liaisons NH et OH via la formation de liaisons hydrogène tandis que les cations se fixent quant à eux principalement sur les sites riches en électrons.

## ***1.2.3. Interactions non-covalentes et capture électronique***

### **1.2.3.1. Les radiations ionisantes**

Il existe différentes sources de radiations qui peuvent être naturelles ou créées par l'homme. L'homme est exposé naturellement à des radiations ionisantes comme celles provenant des rayonnements cosmiques (soleil), mais leurs effets sont atténués grâce à la couche d'ozone. Les sources artificielles se retrouvent dans le domaine médical, la radiologie médicale ou la radiothérapie, dans l'industrie, dans la recherche et le domaine militaire. L'ensemble de ces sources n'est pas forcément bien contrôlé par l'homme et cela peut avoir de graves conséquences pour la santé de l'homme, notamment lors d'accidents nucléaires. La

pénétration naturelle ou accidentelle de ces rayonnements dans l'organisme peut induire des dégâts irréversibles au niveau cellulaire comme la dégradation de l'ADN.

La dégradation de l'ADN par les radiations ionisantes peut se manifester par la cassure d'un ou plusieurs brins de l'ADN, induisant la formation de radicaux, d'ions voire d'électrons libres qui vont interagir à leur tour avec d'autres biomolécules présentes dans la cellule comme les protéines. Mais les processus mis en jeu lors des irradiations de l'ADN sont encore mal connus. Les interactions non covalentes peuvent avoir un rôle dans les processus de radiation ionisante de l'ADN puisqu'elles peuvent influencer les affinités électroniques des groupes présents. Pour mieux comprendre le rôle de ces interactions, beaucoup d'études ont été faites sur l'attachement d'un électron sur les molécules biologiques.[94,95]

Dans la cellule vivante, l'ADN est toujours environné par des protéines et  $H_2O$ . Une récente étude de Sanche et son équipe sur la capture électronique d'un mélange {AA; GCAT}, (AA=Gly ou Arg), a montré que pour une faible quantité d'acide aminé la fragmentation de GCAT est importante, tandis que pour une forte quantité d'acide aminé, celle-ci diminue considérablement.[96] Dans cet exemple, l'acide aminé « protège » l'ADN des irradiations en étant lui-même irradié. La réponse à la question « Où va l'électron ? » est importante. Les sites accepteurs d'électrons vont dépendre de la structure du système qui elle-même dépend des interactions non covalentes. Donc comprendre l'effet de ces interactions sur les sites de capture de l'électron est essentiel.

## ***1.2.4. Interactions non-covalentes et réactivité : cas des interactions avec $H_2O$***

### **1.2.4.1. $H_2O$ , un relai pour les transferts électroniques**

Les réactions de transferts d'électrons sont nombreuses dans les processus biologiques, comme ceux intervenant pour ré-oxyder la NADH déshydrogénase ou dans le fonctionnement des enzymes impliquées dans la réparation de l'ADN (transfert entre la photolyase et l'ADN[97]). Par exemple, De la Lande et ses collaborateurs ont étudié les mécanismes de couplage électronique entre la protéine méthylamine déshydrogénase qui a des propriétés redox et un groupement amicyanine. Ce couple protéine/amicyanine est représentatif de ceux dont dépend la production d'énergie dans les cellules. La force du couplage électronique est facteur à la fois de la conformation adoptée par la protéine, mais également de la possibilité



pour le transfert d'être « assisté » par une ou plusieurs molécule(s) d'eau sous la forme de ponts hydrogène. De la Lande et ses collaborateurs ont ainsi montré que la solvation a un rôle très important dans le processus de transfert électronique puisqu'elle sert de relais pour l'électron par formation de liaisons hydrogène principalement. Voici quelques travaux se focalisant sur l'identification des chemins par lesquels se réalisent les transferts électroniques;[98-100] et sur l'effet de l'eau sur ces transferts d'électrons.[30,101-103]

#### **I.2.4.2. H<sub>2</sub>O, un relai pour les transferts de proton**

Dans le milieu biologique, l'eau peut servir d'agent conducteur d'ions H<sup>+</sup>. De nombreux exemples de ce processus ont été décrits dans la littérature,[30] La plupart des études théoriques effectuées, en complément ou non d'études expérimentales ont examiné la stabilité relative des formes canoniques et zwitterioniques complexées par quelques molécules de H<sub>2</sub>O. Il est étonnant de constater, à ma connaissance, que peu d'études portant sur les barrières de transfert de proton associées au passage entre les deux formes des acides aminés ont été faites à ce jour.[104-108]

De plus, ce type de transfert peut intervenir dans la synthèse des molécules primordiales ou prébiotiques. En effet, après avoir découvert des molécules biologiques telles que les acides aminés dans des météorites et dans les poussières des comètes, l'hypothèse que la vie sur Terre serait d'origine extra-terrestre a été mise en avant. Même si elle est controversée, cela n'empêche que la synthèse de ces systèmes est plausible dans le milieu interstellaire puisqu'ils ont été détectés dans les météorites et comètes. Les glaces d'eau sont très abondantes dans le milieu interstellaire et les comètes, et pourraient participer à la synthèse prébiotique des acides aminés à partir de précurseurs élémentaires éventuellement via des transferts de proton ou d'électrons et (ou) après activation ou via une réaction chimique.

### **I.3. Spectrométrie de masse**

Un large panel de méthodes expérimentales permet de déterminer la structure, certaines propriétés physico-chimiques des peptides et de caractériser les interactions non-covalentes. Dans cette thèse, notre étude sur les interactions non-covalentes a concerné uniquement les molécules formées en phase gazeuse obtenues par la spectrométrie de masse.

Il existe deux types de spectrométrie de masse : « conventionnelle » (Figure I.8) et « en tandem » (Figure I.9).

### Spectrométrie de masse « conventionnelle »

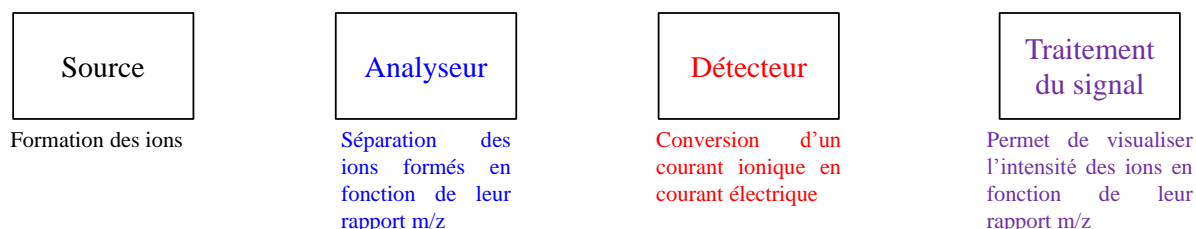


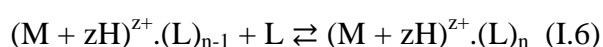
Figure I.8. Spectrométrie de masse conventionnelle

Elle permet d'analyser les ions formés dans la source. Différents modes d'ionisation existent mais nous ne présenterons que les deux sources adaptées à l'étude des peptides. Ces deux sources dites « d'ionisation douce » sont : l'ionisation par électronébulisation (Electrospray Ionisation, ESI) et la Désorption-Ionisation Laser Assistée par Matrice (MALDI). Au cours de l'ESI, les ions préformés en solution sont progressivement désolvatés. Dans le MALDI, il y a des processus de désorption et d'ionisation de l'échantillon à analyser qui a été mis en présence d'une matrice sous l'irradiation laser. Les ions formés dépendent de la matrice utilisée. Ces deux techniques d'ionisation permettent d'analyser par spectrométrie de masse des composés dans une grande gamme de poids moléculaires.

Il existe différents types d'analyseurs : Quadripôle, piège ionique, temps de vol, l'analyseur à secteur magnétique, le FT-ICR (Fourier transform ion cyclotron resonance), l'orbitrap qui se différencient par leur principe de mesure du rapport m/z.

La spectrométrie de masse conventionnelle permet aussi d'accéder à des données thermochimiques comme des constantes de solvation.

L'étude de réactions d'échange de ligands (équation I.6) permet de remonter à ces données thermochimiques.



La constante d'équilibre associée à la réaction (I.6) est donnée par :

$$K_n = \frac{[(M + zH)^{z+} \cdot (L)_n]}{[(M + zH)^{z+} \cdot (L)_{n-1}] \cdot P(L)} \quad (I.7)$$

où  $P(L)$  est la pression partielle du ligand mesurée avec un manomètre. Puisque les intensités des pics  $I_{n-1}$  et  $I_n$  correspondant au poids moléculaire de  $(M + zH)^{z+} \cdot (L)_{n-1}$  et  $(M + zH)^{z+} \cdot (L)_n$  respectivement sont directement proportionnelles à leurs concentrations, alors la constante d'équilibre s'écrit également sous cette forme :

$$K_n = \frac{I_n}{I_{n-1} \cdot P(L)} \quad (I.8)$$

L'énergie libre de Gibbs est directement reliée à la constante d'équilibre par la relation :

$$\Delta G_{n,T}^\circ = -RT \ln K_n \quad (I.9)$$

En outre, si  $K_n$  peut être déterminée à différente température,  $\Delta H_n^\circ$  et  $\Delta S_n^\circ$  sont obtenus par la relation de van't Hoff et sont déterminés expérimentalement en traçant  $\ln K_n$  en fonction de  $\frac{1}{T}$  :

$$\ln K_n = -\frac{\Delta H_{n,T}}{RT} + \frac{\Delta S_{n,T}}{R} \quad (I.10)$$

Avec cette méthode d'équilibre, il est possible de mesurer les données thermochimiques d'hydratation successive.

### Spectrométrie de masse « en tandem »

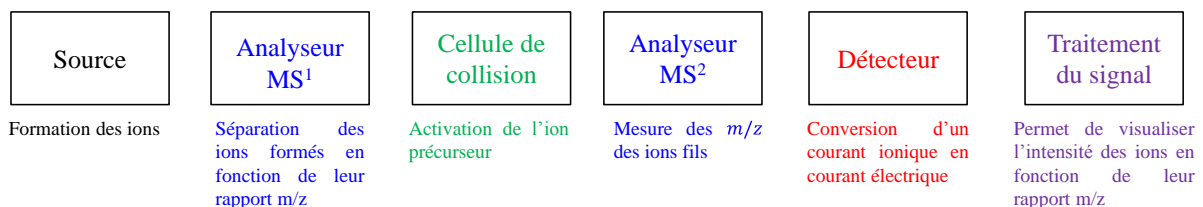
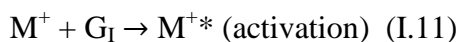


Figure I.9. Spectromètre de masse en tandem (MS/MS)

Elle consiste à isoler un ion formé en source (appelé « précurseur » ou « parent ») et à l'introduire dans une cellule de collision où se trouve un gaz de collision inerte ( $G_I$ ) dans le cas de la dissociation induite par collision (CID), la plus utilisée. Les collisions induisent un transfert de l'énergie cinétique de l'ion en énergie interne (excitation vibrationnelle).

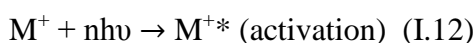


L'ion activé  $M^{+*}$  peut alors se fragmenter au niveau des liaisons les plus faibles en éliminant une molécule neutre. Ce type d'analyse permet d'obtenir des informations structurales sur la molécule étudiée par l'analyse des ions fragments.

D'autres modes d'activations mettant en jeu des processus d'excitations vibrationnelles (IRMPD) ou électroniques (Dissociation Induite par les Électrons (EID)) ont été utilisés pour la caractérisation de peptides [109].

### **Dissociation multiphotonique**

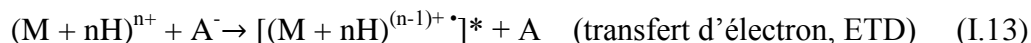
Dans ce type de dissociation, l'ion précurseur est excité par l'absorption d'un ou plusieurs photons infra-rouge (IR). Ces photons peuvent être issus d'un laser IR (à  $CO_2$ ,  $\nu = 944 \text{ cm}^{-1}$  par exemple) voire d'un corps noir (radiation thermique). Dans le premier cas, on parlera de Dissociation MultiPhotonique IR (IRMPD) et dans le deuxième cas, de Dissociation IR de corps noir (BIRD).



Pour qu'il y ait fragmentation, l'énergie absorbée doit être supérieure à l'énergie de dissociation de la voie dissociative la plus favorable.

### **Dissociation induite par des électrons[110]**

Ce troisième mode d'activation couramment utilisé consiste à réduire un ion précurseur multichargé positivement par absorption d'un électron. Il existe deux grandes méthodes :



Dans le premier cas, un transfert d'électron est réalisé par une réaction (ion-molécule) entre un anion réactif  $A^-$  et l'espèce cationique multichargée  $(M + nH)^{n+}$ . Dans le deuxième cas, l'électron capturé par l'ion multichargé provient d'un faisceau d'électrons. Dans les deux cas le processus est exothermique, par conséquent le radical cation  $[M + nH]^{(n-1)+\bullet}$  formé va avoir un excédent d'énergie interne qui entraîne sa fragmentation.

## I.4. Apport de la chimie théorique

Avec la spectrométrie de masse, l'étude des masses obtenues, des fragments et des données thermochimiques permet d'obtenir des propriétés physico-chimiques des biomolécules. À partir des fragments, il est plus facile de déterminer la structure si les mécanismes de fragmentations sont bien connus, ce qui n'est pas toujours le cas et empêche une détermination automatique de la structure primaire. De plus, par le biais des données thermochimiques mesurées expérimentalement, il est possible de justifier les tendances observées en comparant plusieurs systèmes de même nature ou qui possède un groupement en commun. Mais identifier tous les conformères contribuant à ces données expérimentales s'avère être de plus en plus difficile lorsque la taille de la biomolécule étudiée devient importante. Pour remédier à cette limite expérimentale, les outils de la modélisation moléculaire et de la chimie quantique sont mises à contributions pour permettre de mieux comprendre les observations faites expérimentalement allant jusqu'à parfois les prédire.

Dans le cadre de notre étude, une bonne description des interactions non-covalentes au sein des systèmes passe d'abord par une élucidation structurale. En fonction de la taille et de la flexibilité du système, la recherche conformationnelle (manuelle ou non) se fait couramment avec des « bas niveaux » de théorie comme l'utilisation de champ de forces, méthodes semi-empiriques, HF ou DFT. Les grandeurs caractéristiques des systèmes (thermochimiques, énergies d'interactions, ...) sont plutôt calculées à partir de niveaux de théorie plus élevées (méthodes composites, post-HF par exemple) pour qu'elles soient plus précises qu'en DFT. Dans ce travail, les explorations de surfaces d'énergie potentielle, les caractéristiques structurales, thermochimiques et électroniques ont nécessité de trouver les bonnes méthodes de calculs et stratégies pour décrire aux mieux les effets des interactions non-covalentes. Par conséquent, nous avons utilisé un large éventail de méthodes de calculs pour les adapter aux systèmes et à ce que nous avons voulu mettre en évidence. Dans le chapitre suivant, nous présentons les méthodes, outils et stratégies utilisées dans nos travaux.

## I.5. Bibliographie

- [1] Israelachvili, J. *Intermolecular & Surfaces Forces*, San Diego, *Academic Press*, **1992**, 450 pages.
- [2] Granger, P. *Panorama des liaisons chimiques*, Paris, *Elsevier-Masson*, **1997**, 198 pages.
- [3] Gerschel, A. *Liaisons intermoléculaires – Les forces en jeu dans la matière condensée*, Paris, *InterEditions/CNRS Editions*, **1995**, 274 pages.
- [4] Dougherty, D. A. *J. Nutr.* **2007**, *137*, 1504S-1508S.
- [5] Tommos, C.; Babcock, G. T. *Biochim. Biophys. Acta* **2000**, *1458*, 199-219.
- [6] Hosler, J. P.; Ferguson-Miller, S.; Mills, D. A. *Annu. Rev. Biochem.* **2006**, *75*, 165-187.
- [7] Harrison, A. G. *Mass Spectrom. Rev.* **1997**, *16*, 201-217.
- [8] Bouchoux, G. *Mass Spectrom. Rev.* **2014**, DOI: 10.1002/mas.21416.
- [9] Bouchoux, G. *Mass Spectrom. Rev.* **2012**, *31*, 391-435.
- [10] Riffet, V.; Frison, G.; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18561-18580.
- [11] Riffet, V.; Bourcier, S.; Bouchoux, G. *Int. J. Mass Spectrom.* **2012**, *316-318*, 47-56.
- [12] Riffet, V.; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6097-6106.
- [13] Uddin, K. M.; Warburton, P. L.; Poirier, R. *J. Phys. Chem. B* **2012**, *116*, 3220-3234.
- [14] Gronert, S.; Simpson, D. C.; Conner, K. M. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2116-2123.
- [15] Auffinger, P.; Hayes, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16789-16794.
- [16] Scholfield, M. R.; Vander Zanden, C. M.; Carter, M.; Ho, P. S. *Protein Sci.* **2013**, *22*, 139-152.
- [17] Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G. *Acc. Chem. Res.* **2005**, *38*, 386-395.
- [18] Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974-12975.
- [19] Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 3494-3501.
- [20] Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2004**, *108*, 4885-4898.
- [21] Sudha, R.; Kohtani, M.; Jarrold, M. F. *J. Phys. Chem. B* **2005**, *109*, 6442-6447.
- [22] Kinnear, B. S.; Hartings, M. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **2001**, *123*, 5660-5667.
- [23] Kinnear, B. S.; Kaleta, D. T.; Kohtani, M.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 9243-9256.
- [24] Kaleta, D. T.; Jarrold, M. F. *J. Phys. Chem. B* **2001**, *105*, 4436-4440.
- [25] Dean, J. C.; Buchanan, E. G.; Zwier, T. S. *J. Am. Chem. Soc.* **2012**, *134*, 186-201.
- [26] Martens, J. K.; Compagnon, I.; Nicol, E.; McMahon, T. B.; Clavaguéra, C.; Ohanessian, G. *J. Phys. Chem. Lett.* **2012**, *3*, 3320-3324.
- [27] Tkatchenko, A.; Rossi, M.; Blum, V.; Ireta, J.; Scheffler, M. *Phys. Rev. Lett.* **2011**, *106*, 118102-1-4.
- [28] Rossi, M.; Blum, V.; Kupser, P.; vonHelden, G.; Bierau, F.; Pagel, K.; Meijer, G.; Scheffler, M. *J. Phys. Chem. Lett.* **2010**, *1*, 3465-3470.
- [29] Kohtani, M.; Schneider, J. E.; Jones, T. C.; Jarrold, M. F. *J. Am. Chem. Soc.* **2004**, *126*, 16981-16987.
- [30] Ball, P. *Chem. Rev.* **2008**, *108*, 74-108.
- [31] Meot-Ner, M. *Chem. Rev.* **2005**, *105*, 213-284.
- [32] Wyttenbach, T.; Bowers, M. T. *Chem. Phys. Lett.* **2009**, *480*, 1-16.
- [33] Meng, X. J.; Zhao, H. L.; Ju, X. S. *Comput. Theor. Chem.* **2012**, *1001*, 26-32.
- [34] Balta, B.; Aviyente, V. *J. Comput. Chem.* **2003**, *24*, 1789-1802.
- [34] Tiwari, S.; Mishra, P. C.; Suhai, S. *Int. J. Quantum Chem.* **2008**, *108*, 1004-1016.
- [35] Kim, J. Y.; Im, S.; Kim, B.; Desfrancois, C.; Lee, S. *Chem. Phys. Lett.* **2008**, *451*, 198-203.

- [36] Pluhařová, E.; Marsalek, O.; Schmidt, B.; Jungwirth, P. *J. Chem. Phys.* **2012**, *137*, 185101-1-8.
- [37] Miyamoto, H.; Aida, M. *Chem. Lett.* **2013**, *42*, 598-600.
- [38] Kim, J. Y.; Ahn, D. S.; Park, S. W.; Lee, S. *RSC Adv.* **2014**, *4*, 16352-16361.
- [39] Mercier, S. R.; Boyarkin, O. V.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.; Rothlisberger, U.; Rizzo, T. R. *J. Am. Chem. Soc.* **2006**, *128*, 166938-16943.
- [40] Biswal, H. S.; Loquais, Y.; Tardivel, B.; Gloaguen, E.; Mons, M. *J. Am. Chem. Soc.* **2011**, *133*, 3931-3942.
- [41] Avbelj, F.; Luo, P.; Baldwin, R. L. *Proc. Natl. Acad. USA* **2000**, *97*, 10786-10791.
- [42] Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2002**, *124*, 11148-11158.
- [43] Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2004**, *126*, 8454-8458.
- [44] Chutia, S.; Rossi, M.; Blum, V. *J. Phys. Chem. B* **2012**, *116*, 14788-14804.
- [45] Levy, Y.; Onuchic, J. N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389-415.
- [46] Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. *J. Chem. Inf. Model.* **2007**, *47*, 668-675.
- [47] Wu, R.; McMahon, T. B. *Can. J. Chem.* **2005**, *83*, 1978-1993.
- [48] Wu, R.; McMahon, T. B. *J. Am. Chem. Soc.* **2008**, *130*, 3065-3078.
- [49] Wu, R.; McMahon, T. B. *J. Mass Spectrom.* **2008**, *43*, 1641-1648.
- [50] Wu, R.; McMahon, T. B. *Angew. Chem. Int. Ed.* **2007**, *46*, 3668-3671.
- [51] Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M. *J. Biol. Inorg. Chem.* **2008**, *13*, 1205-1218.
- [52] Lu, Y. *Proc. Natl. Acad. Sci.* **2010**, *107*, 1811-1812.
- [53] Holm, H. R.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239-2314.
- [54] Hoyau, S.; Norrman, K.; McMahon, T. B.; Ohanessian, G. *J. Am. Chem. Soc.* **1999**, *121*, 8864-8875.
- [55] Talley, J. M.; Cerda, B. A.; Ohanessian, G.; Wesdemiotis, C. *Chem. Eur. J.* **2002**, *8*, 1377-1388.
- [56] Wang, P.; Ohanessian, G.; Wesdemiotis, C. *Eur. J. Mass Spectrom.* **2009**, *15*, 325-335.
- [57] Corral, I.; Mó, O.; Yáñez, M.; Salpin, J. Y.; Tortajada, J.; Moran, D.; Radom, L. *Chem. Eur. J.* **2006**, *12*, 6787-6796.
- [58] Mó, O.; Yáñez, M.; Salpin, J. Y.; Tortajada, J. *Mass Spectrom. Rev.* **2007**, *26*, 474-516.
- [59] Ai, H.; Bu, Y.; Han, K. *J. Chem. Phys.* **2003**, *118*, 10973-10985.
- [60] Remko, M.; Rode, B. M. *J. Phys. Chem. A* **2006**, *110*, 1960-1967.
- [61] Remko, M.; Fitz, D.; Rode, B. M. *Amino acids* **2010**, *39*, 1309-1319.
- [62] Remko, M.; Fitz, D.; Rode, B. M. *J. Phys. Chem. A* **2008**, *112*, 7652-7661.
- [63] Bush, M. F.; Oomens, J.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2008**, *130*, 6463-6471.
- [64] O'Brien, J. T.; Prell, J. S.; Steill, J. D.; Oomens, J.; Williams, E. R. *J. Phys. Chem. A* **2008**, *112*, 10823-10830.
- [65] Dunbar, R. C.; Hopkinson, A. C.; Oomens, J.; Siu, C. K.; Siu, K. W. M.; Steill, J. D.; Verkerk, U. H.; Zhao, J. *J. Phys. Chem. B* **2009**, *113*, 10403-10408.
- [66] Remko, M.; Fitz, D.; Broer, R.; Rode, B. M. *J. Mol. Model.* **2011**, *17*, 3117-3128.
- [67] Remko, M.; Šoralová, S. *J. Biol. Inorg. Chem.* **2012**, *17*, 621-630.
- [68] Armentrout, P. B.; Rodgers, M. T.; Oomens, J.; Steill, J. D. *J. Phys. Chem. A* **2008**, *112*, 2248-2257.
- [69] Armentrout, P. B.; Yang, B.; Rodgers, M. T. *J. Phys. Chem. B* **2013**, *117*, 3771-3781.
- [70] Hurtado, M.; Monte, M.; Lamsabhi, A. M.; Yáñez, M.; Mó, O.; Salpin, J. Y. *ChemPlusChem* **2013**, *78*, 1124-1133.
- [71] Larrucea, J.; Rezabal, E.; Marino, T.; Russo, N.; Ugalde, J. M. *J. Phys. Chem. B* **2010**, *114*, 9017-9022.

- [72] a) Burt, M. B.; Decker, S. G. A.; Atkins, C. G.; Rowsell, M.; Peremans, A.; Fridgen, T. D. *J. Phys. Chem. B* **2011**, *115*, 11506-11518. b) Burt, M. B.; Fridgen, T. D. *J. Phys. Chem. A* **2013**, *117*, 1283-1290.
- [73] Burford, N.; Eelman, M. D.; Groom, K. *J. Inorg. Biochem.* **2005**, *99*, 1992-1997.
- [74] Banu, L.; Blagojevic, V.; Bohme, D. K. *Int. J. Mass Spectrom.* **2012**, *330-332*, 168-173.
- [75] Atkins, C. G.; Banu, L.; Rowsell, M.; Blagojevic, V.; Bohme, D. K.; Fridgen, T. D. *J. Phys. Chem. B* **2009**, *113*, 14457-14464.
- [76] Zatta, P.; Lucchini, R.; van Rensburg, S. J.; Taylor, A. *Brain Res. Bull.* **2003**, *62*, 15-28.
- [77] Kiss, T.; Gajda-Schranz, K.; Zatta, P. F. The Role of Aluminum in Neurotoxic and Neurodegenerative Processes, in Neurodegenerative Diseases and Metal Ions, Volume 1 (eds A. Sigel, H. Sigel and R. K. O. Sigel), Chichester, *John Wiley & Sons*, **2006**, pages 488.
- [78] Latrous, L.; tortajadaa, J.; Haldys, V.; Léon, E.; Correia, C.; salpin, J. Y. *J. Mass Spectrom.* **2013**, *48*, 795-806.
- [79] Lemoff, A. S.; Bush, M. F.; O'Brien, J. T.; Williams, E. R. *J. Phys. Chem. A* **2006**, *110*, 8433-8442.
- [80] Lemoff, A. S.; Wu, C. C.; Bush, M. F.; Williams, E. R. *J. Phys. Chem. A* **2006**, *110*, 3662-3669.
- [81] Lemoff, A. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1014-1024.
- [82] Jockusch, R. A.; Lemoff, A. S.; Williams, E. R. *J. Phys. Chem. A* **2001**, *105*, 10929-10942.
- [83] Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 13576-13584.
- [84] Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Phys. Chem. A* **2005**, *109*, 1903-1910.
- [85] Qin, P. H.; Zhang, W.; Lu, W. C. *Comput. Theor. Chem.* **2013**, *1021*, 164-170.
- [86] Gutowski, M.; Skurski, P.; Simons, J. *J. Am. Chem. Soc.* **2000**, *112*, 10159-10162.
- [87] Milner, E. M.; Nix, M. G. D.; Dessent, C. E. H. *J. Phys. Chem. A* **2012**, *116*, 801-809.
- [88] O'Brien, J. T.; Prell, J. S.; Berden, G.; Oomens, J.; Williams, E. R. *Int. J. Mass Spectrom.* **2010**, *297*, 116-123.
- [89] Schmidt, J.; Kass, S. R. *J. Phys. Chem. A* **2013**, *117*, 4863-4869.
- [90] Walter, M.; Sen, A.; Harvey, A. J. A.; Dessent, C. E. H. *Chem. Phys. Lett.* **2013**, *588*, 43-46.
- [91] Kass, S. R. *J. Am. Chem. Soc.* **2005**, *127*, 13098-13099.
- [92] Yang, G.; Zu, Y.; Liu, C.; Fu, Y.; Zhou, L. *J. Phys. Chem. B* **2008**, *112*, 7104-7110.
- [93] Tian, S. X.; Li, H. B.; Yang, J. *ChemPhysChem* **2009**, *10*, 1435-1437.
- [94] Alizadeh, E.; Sanche, L. *Chem. Rev.* **2012**, *112*, 5578-5602.
- [95] Gu, J.; Leszczynski, J.; Schaefer III, H. F. *Chem. Rev.* **2012**, *112*, 5603-5640.
- [96] Ptasíńska, S.; Li, Z.; Mason, N. J.; Sanche, L. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9367-9372.
- [97] Sancar, A. *Chem. Rev.* **2003**, *103*, 2203-2237.
- [98] De la Lande, A.; Babcock, N. S.; Řezáč, J.; Sanders, B. C.; Salahub, D. R. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11799-11804.
- [99] De la Lande, A.; Babcock, N. S.; Řezáč, J.; Lévy, B.; Sanders, B. C.; Salahub, D. R. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5902-5918.
- [100] Lin, J.; Baladin, I. A.; Beratan, D. N. *Science* **2005**, *310*, 1311-1313.
- [101] Beratan, D. N.; Skourtis, S. S.; Balabin, I. A.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Xiao, D. *Acc. Chem. Res.* **2009**, *42*, 1669-1678.
- [102] Baladin, I. A.; Beratan, D. N. *Phys. Rev. Lett.* **2008**, *101*, 158102-1-4.
- [103] Migliore, A.; Corni, S.; Di Felice, R.; Molinari, E. *J. Phys. Chem. B* **2007**, *111*, 3774-3781.
- [104] Rodziewicz, P.; Doltsinis, N. L. *ChemPhysChem* **2007**, *8*, 1959-1968.



- [105] Kim, J. Y.; Im, S.; Kim, B.; Desfrancois, C.; Lee, S. *Chem. Phys. Lett.* **2008**, *451*, 198-203.
- [106] Tiwari, S.; Mishra, P. C.; Suhai, S. *Int. J. Quantum Chem.* **2008**, *108*, 1004-1016.
- [107] Tian, S. X.; Sun, X.; Cao, R.; Yang, J. *J. Phys. Chem. A* **2009**, *113*, 480-483.
- [108] Meng, X. J.; Zhao, H. L.; Ju, X. S. *Comput. Theor. Chem.* **2012**, *1001*, 26-32.
- [109] Sleno, L.; Volmer, D. A. *J. Mass Spectrom.* **2004**, *39*, 1091-1112.
- [110] Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. *Chem. Soc. Rev.* **2013**, *42*, 5014-5030.

## Chapitre II : Méthodologies



Cette session a été inspirée de divers livres de chimie quantique et de mécanique moléculaire,[1-3] mais aussi d'articles scientifiques qui seront mentionnés dans le texte. Nous y présentons les méthodes et les outils utilisés au cours de ces travaux de thèse.

## II.1. Les méthodes *ab initio*

Contrairement à la mécanique classique, la mécanique quantique est probabiliste ; en effet selon le principe d'incertitude d'Heisenberg, il est impossible de déterminer simultanément la position et la vitesse de la particule à un instant  $t$  donné. Toutefois, il est possible de calculer la probabilité  $P(r,t)$  pour qu'une particule se trouve en une position  $r$  et à un temps  $t$  donné. Elle correspond à  $\Psi^2(r,t)$ , où  $\Psi(r,t)$  est la fonction d'onde obtenue par la résolution de l'équation de Schrödinger dépendante du temps pour une particule non-relativiste :

$$\hat{H}(r,t)\Psi(r,t) = i \frac{\partial \Psi(r,t)}{\partial t} \quad (\text{II.1.1})$$

$\hat{H}(r,t)$  est l'opérateur hamiltonien du système dépendant du temps.

En régime stationnaire, l'équation de Schrödinger devient :

$$\hat{H}\Psi(r) = E\Psi(r) \quad (\text{II.1.2})$$

où  $E$  correspond à l'énergie du système dans son état stationnaire.

L'opérateur hamiltonien poly-électronique s'écrit comme la somme des opérateurs énergie cinétique  $\hat{T}$  et énergie potentielle  $\hat{V}$  :

$$\hat{H} = \hat{T} + \hat{V} \quad (\text{II.1.3})$$

avec

$$\hat{T} = -\sum_{k=1}^M \frac{\nabla_k^2}{2M_k} - \sum_{i=1}^n \frac{\nabla_i^2}{2} \quad (\text{II.1.4})$$

représentant la somme de l'énergie cinétique de  $M$  noyaux de masse  $M_k$  et de  $n$  électrons, et

$$\hat{V} = -\sum_{i=1}^n \sum_{k=1}^M \frac{Z_k}{R_{i,k}} + \sum_{i=1}^n \sum_{j<i} \frac{1}{r_{ij}} + \sum_{k=1}^M \sum_{l<k} \frac{Z_k Z_l}{R_{k,l}} \quad (\text{II.1.5})$$

représentant l'attraction coulombienne électron-noyau, ainsi que les répulsions coulombiennes électron-électron et noyau-noyau respectivement.

Excepté pour l'atome d'hydrogène et les ions hydrogénoïdes, la résolution exacte de l'équation de Schrödinger est actuellement impossible à cause du terme biélectronique. Un ensemble d'approximations a été utilisé pour permettre des résolutions approchées de l'équation (II.1.2). Celles utilisées le plus couramment en chimie quantique sont présentées ci-dessous.

### ***II.1.1. La méthode Hartree-Fock***

Dans le cadre de l'approximation adiabatique et de Born-Oppenheimer, le terme spécifique aux noyaux de l'équation (II.1.4) est négligé et la répulsion coulombienne noyaux-noyaux de l'équation (II.1.5) devient constante. Les  $n$  électrons sont considérés indépendamment les uns des autres et évoluent dans un champ moyen créé par les  $n-1$  autres électrons et les  $N$  noyaux fixes. Cela revient à définir pour chaque électron un hamiltonien mono-électronique de cœur  $\hat{H}^c(i)$  par :

$$\hat{H}^c(i) = -\frac{\nabla_i^2}{2} - \sum_{k=1}^M \frac{Z_k}{r_{i,k}}(i) \quad (\text{II.1.6})$$

et par conséquent l'hamiltonien électronique du système devient:

$$\hat{H} = \sum_{i=1}^n \hat{H}^c(i) + \sum_{i=1}^n \sum_{j<i} \frac{1}{r_{ij}} \quad (\text{II.1.7})$$

Avec l'approximation de Born-Oppenheimer, on suppose alors que l'écart énergétique entre les états électroniques est élevé (plusieurs eV de différence).

L'énergie électronique du système est donnée par la relation suivante :

$$E = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \quad (\text{II.1.8})$$

La fonction d'onde  $\Psi(1,2,\dots,n)$  du système, dans le cadre de l'approximation orbitale, dans laquelle on néglige l'interaction électron-électron, et par le respect du principe d'exclusion de Pauli, est un déterminant de Slater construit sur la base de  $n$  spin-orbitales orthonormées  $\chi_i$ :

$$\Psi(1,2,\dots,n) = |\chi_1 \chi_2 \dots \chi_n| \quad (\text{II.1.9})$$

avec

$$\chi_i = \phi_i \eta_i \quad (\text{II.1.10})$$

où  $\eta_i$  est la fonction de spin  $\alpha$  ou  $\beta$  et  $\phi_i$  est la fonction d'espace.

L'énergie du système donné à l'équation (II.1.8) devient :

$$E = \sum_{i=1}^n \langle \chi_i(1) | \hat{H}^c(1) | \chi_i(1) \rangle + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\hat{J}_{ij} - \hat{K}_{ij}) \quad (\text{II.1.11})$$

$\hat{J}_{ij}$  est l'intégrale coulombienne biélectronique

$$\hat{J}_{ij} = \langle \chi_i(1) \chi_j(2) | \frac{1}{r_{12}} | \chi_i(1) \chi_j(2) \rangle \quad (\text{II.1.12})$$

$\hat{K}_{ij}$  est l'intégrale d'échange biélectronique

$$\hat{K}_{ij} = \langle \chi_i(1) \chi_j(2) | \frac{1}{r_{12}} | \chi_j(1) \chi_i(2) \rangle \quad (\text{II.1.13})$$

Selon la méthode variationnelle, la meilleure estimation de l'énergie est celle obtenue par la fonction d'onde  $|\Psi\rangle$  qui minimise l'énergie  $E$ . Cette condition revient à résoudre un ensemble de  $n$  équations d'un opérateur mono-électronique, à savoir l'opérateur de Fock:

$$\hat{f}_i \chi_i = \varepsilon_i \chi_i \quad (\text{II.1.14})$$

avec

$$\hat{f}_i = \hat{H}^c(1) + \sum_{j=1}^n (\hat{J}_j - \hat{K}_j) \quad (\text{II.1.15})$$

où  $\hat{J}_j$  et  $\hat{K}_j$  sont l'opérateur de Coulomb et d'échange respectivement.

En utilisant l'approche LCAO (Linear Combinaison of Atomic Orbitals) (II.1.16) et les équations de Roothaan-Hall, la résolution de l'ensemble de ces  $n$  équations de Fock est

obtenue par un processus itératif nommé méthode du champ autocohérent (Self Consistent Field (SCF)).

$$\phi_i = \sum_{\alpha} c_{\alpha i} \varphi_{\alpha} \quad (\text{II.1.16})$$

$\phi_i$  est une combinaison linéaire finie d'orbitales atomiques.

Deux points importants doivent être notés ici. D'abord, l'interaction coulombienne d'un électron avec lui-même (ou self interaction) se compense exactement avec l'interaction d'échange car  $J_{ii} = K_{ii}$ . Cette compensation d'erreur ne se retrouvera pas dans toutes les méthodes utilisées comme nous le verrons par la suite. Puis, dans l'approximation Hartree-Fock, la probabilité que deux électrons de spin opposés soient à la même position dans l'espace est non nulle. Cette erreur se nomme erreur de corrélation. En réalité, les électrons n'étant pas indépendants les uns des autres, la probabilité que deux électrons soient très proches dans l'espace est surestimée dans le cadre de Hartree-Fock. L'énergie de corrélation  $E_c$  est définie de la façon suivante :

$$E_c = E_0 - E_{HF} \quad (\text{II.1.17})$$

avec  $E_0$  l'énergie exacte du système. L'énergie de corrélation peut être calculée par différentes méthodes de calculs que nous verrons également par la suite.

Dans le cadre de la méthode Hartree-Fock, il est possible d'imposer des contraintes de spin (Restricted Hartree-Fock (RHF)) ou non (Unrestricted Hartree-Fock (UHF)). Précédemment, nous nous étions placés dans le cas RHF, c'est-à-dire que la fonction d'espace  $\phi_i$  est la même pour les deux spin-orbitales  $\chi_i^{\alpha}$  et  $\chi_i^{\beta}$  :

$$\chi_i^{\alpha} = \phi_i \alpha_i \quad (\text{II.1.18})$$

$$\chi_i^{\beta} = \phi_i \beta_i \quad (\text{II.1.19})$$

Dans le formalisme UHF, les spins  $\alpha$  et  $\beta$  peuvent avoir des fonctions d'espace  $\phi_i$  différentes et donc des énergies différentes. Ainsi, les spin-orbitales ont pour expressions :

$$\chi_i^{\alpha} = \phi_i^{\alpha} \alpha_i \quad (\text{II.1.20})$$

$$\chi_i^{\beta} = \phi_i^{\beta} \beta_i \quad (\text{II.1.21})$$

La fonction d'onde UHF  $\Psi_{UHF}$  n'est pas une fonction propre de l'opérateur  $S^2$ , sauf si  $\Psi_{UHF} = \Psi_{RHF}$ . Dans ce dernier cas, la valeur propre attendue pour cet opérateur est  $\langle S^2 \rangle = S_z(S_z + 1)$ . Sinon, cette valeur est calculée par l'équation suivante :

$$\langle S^2 \rangle = S_z(S_z + 1) + N_\beta - \sum_{ij} \langle \chi_i^\alpha | \chi_j^\beta \rangle^2 \quad (\text{II.1.22})$$

avec  $N_\beta$  le nombre de spin-orbitales  $\beta$ . La somme des deux derniers termes, terme toujours positif, représente la contamination de spin. Plus  $\Psi_{UHF}$  est proche de  $\Psi_{RHF}$ , plus la contamination de spin sera faible. Dans ce travail de thèse, pour l'étude des systèmes à couches ouvertes, nous nous placerons dans le cadre « Unrestricted ».

## II.1.2. Les méthodes post Hartree-Fock

Ces méthodes vont permettre de prendre en compte la corrélation des mouvements électroniques. On définit deux types de corrélation électronique : la corrélation statique, qui provient de la nature multiconfigurationnelle de la fonction d'onde, et la corrélation dynamique, relative à la probabilité non nulle d'avoir 2 électrons au même endroit.

### II.1.2.1. Interaction de configurations

La méthode d'interaction de configurations (CI) est basée sur le principe variationnel. La fonction d'onde  $\Psi_{CI}$  est construite à partir d'une combinaison linéaire de fonctions d'état de configurations, résultant elles-mêmes d'une combinaison linéaire de déterminants de Slater, de l'état fondamental et des états excités. Les configurations excitées sont construites sur la base de  $\Psi_{HF}$  et sont obtenues en excitant les électrons vers les orbitales virtuelles.

$$\Psi_{CI} = c_0 \Psi_{HF} + \sum_i \sum_a C_i^a \Psi_i^a + \sum_{ij} \sum_{ab} C_{ij}^{ab} \Psi_{ij}^{ab} + \sum_{ijk} \sum_{abc} C_{ijk}^{abc} \Psi_{ijk}^{abc} \quad (\text{II.1.23})$$

Le deuxième, troisième et quatrième termes de l'équation (II.1.23) indiquent que les déterminants de Slater correspondent à des états simplement, doublement et triplement excités par rapport au déterminant HF, respectivement.

Les coefficients  $C_{i(i=S,D,T,...)}$  sont optimisés suivant le principe variationnel. Cette fonction permet d'atteindre une valeur exacte de  $E_0$  si toutes les excitations multiples possibles à partir



de  $\Psi_{HF}$  sont prises en compte puisque dans  $\Psi_{CI}$  sont incluses toutes les permutations des  $n$  électrons dans toutes les orbitales virtuelles. Ce procédé se nomme full-CI et est très coûteux en temps de calcul. En fonction du système à étudier, certains types d'excitation peuvent être ignorés pour simplifier le calcul, comme les multiples excitations dans la méthode CIS. Il existe une extension de l'interaction de configuration, il s'agit de l'interaction de configurations quadratiques QCISD.[4] Les équations sont modifiées en ajoutant de nouveaux termes quadratiques dans les coefficients de configurations. Cela permet de corriger les erreurs de cohérence qui sont dues à la taille du système étudié.

### II.1.2.2. Le champ multi-configurationnel auto-cohérent (MCSCF)

La méthode d'interaction de configurations est utilisée lorsque la contribution de  $\Psi_{HF}$  est prédominante, on parle alors de systèmes mono-configurationnels. Toutefois, lorsque l'écart en énergie entre les OM occupées et vacantes devient faible, la contribution de  $\Psi_{HF}$  n'est plus prédominante. On dit alors que le système est multi-configurationnel. Ainsi, en plus de l'optimisation des coefficients d'interaction de configuration  $C_{i(i=S,D,T,...)}$  par la méthode variationnelle, les coefficients des orbitales atomiques  $c_{\alpha i}$  dans les déterminants de Slater sont optimisés simultanément. Ces optimisations MCSCF sont itératives comme dans le cas de la procédure classique SCF et sont par conséquent plus longues en temps de calcul et plus gourmandes en mémoire.

### II.1.2.3. Complete Active Space Self-Consistent Field (CASSCF)

La méthode CASSCF est un cas particulier de la méthode MCSCF qui permet d'obtenir une bonne description de l'état fondamental et des états excités. Elle consiste à partitionner l'ensemble des OM en trois espaces distincts : un espace inactif (qui contient des OM doublement occupées optimisées dans le calcul itératif), un espace actif et un espace virtuel (qui contient des orbitales vides). Un calcul full-CI est réalisé dans l'espace actif et les OM obtenues ont une occupation qui varie entre 0 et 2. Le choix de cet espace est crucial afin de bien prendre en compte les orbitales impliquées dans le phénomène chimique que l'on souhaite étudier. Les OM actives sont souvent les OM les plus hautes occupées et les plus basses vacantes. On choisit ainsi  $m$  OM et  $n$  électrons du système étudié pour construire l'espace actif qui sera noté CASSCF( $n,m$ ).

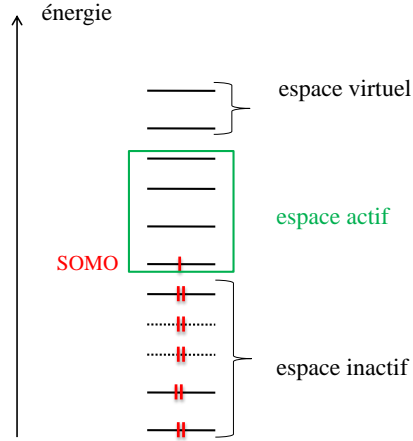


Figure II.1. Diagramme des OM d'un système fictif radicalaire et exemple de partitionnement des OMs.

#### II.1.2.4. Interaction de configurations multi-référence (MRCI)

La méthode MRCI est une méthode d'interaction de configurations pour laquelle la fonction d'onde est multi-configurationnelle. La fonction d'onde  $\Psi_{MCSCF}$  est utilisée comme référence. Le calcul MRCI sert à optimiser davantage la fonction d'onde de chaque état provenant de  $\Psi_{MCSCF}$ . Un calcul CI est réalisé en considérant toutes les simples et doubles excitations de ces configurations de référence.

La fonction d'onde MRCI s'écrit de la manière suivante :

$$\Psi_{MRCI} = a_0 \Psi_{MCSCF} + \sum_i \sum_a C_i^a \Psi_i^a + \sum_{ij} \sum_{ab} C_{ij}^{ab} \Psi_{ij}^{ab} \quad (\text{II.1.24})$$

avec  $\Psi_S^a$  et  $\Psi_D^{ab}$  les fonctions de configuration obtenues à partir de la fonction  $\Psi_{MCSCF}$  dans le cas d'une simple et d'une double excitation, respectivement.

La correction de Davidson permet de corriger les erreurs de cohérence de taille en prenant en compte les excitations quadripolaires par l'ajout d'un terme correctif à l'énergie :

$$\Delta E_Q = (1 - C_0^2) \Delta E_{CISD} \quad (\text{II.1.25})$$

avec  $C_0$  le coefficient d'interaction de configurations de  $\Psi_{HF}$ .

### II.1.2.5. Coupled Cluster (CC)

La méthode « Coupled Cluster » fait partie des méthodes les plus précises pour décrire la fonction d'onde à un déterminant de l'état fondamental qui s'écrit :

$$\Psi_{CC} = e^{\hat{T}} \Psi_{HF} \quad (\text{II.1.26})$$

avec  $\hat{T}$  l'opérateur agissant sur  $\Psi_{HF}$  qui génère toutes les excitations des électrons des orbitales occupées vers les orbitales vacantes ayant comme expression :

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \quad (\text{II.1.27})$$

$\hat{T}_1$  est l'opérateur de toutes les mono-excitations,  $\hat{T}_2$  des doubles excitations, ....

La méthode CCSD prend en compte seulement les mono et doubles excitations, l'opérateur  $\hat{T}$  étant de la forme :

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \quad (\text{II.1.28})$$

L'ajout des triples excitations par la méthode des perturbations donne la méthode CCSD(T) qui fait référence actuellement en chimie quantique.

La méthode EOM-CC (*Equation of motion coupled-cluster*) peut être considérée comme une extension de la méthode Coupled Cluster qui permet de donner des solutions approchées des états excités du système étudié.[5] Cette méthode est couteuse en temps de calcul mais est souvent utilisée pour valider des résultats TD-DFT sans avoir besoin de recourir à des méthodes multi-configurationnelles.[6]

Nous avons utilisé les méthodes EOM-CCSD et CASSCF-MRCI comme méthodes de référence pour les données TD-DFT concernant la structure électronique de l'état fondamental et des états excités de peptides radicaux cations (chapitre IV).

### II.1.2.6. La théorie des perturbations : Møller-Plesset

Afin de prendre en compte la corrélation électronique, une autre approche, moins couteuse que les précédentes, basée sur la théorie des perturbations a été proposée. Il s'agit de la théorie de perturbation de Møller-Plesset qui n'est pas variationnelle. L'opérateur hamiltonien du système est défini par :

$$\hat{H} = \hat{H}^{(0)} + \hat{V} \quad (\text{II.1.29})$$

avec  $\hat{H}^{(0)}$  l'hamiltonien à l'ordre zéro et  $\hat{V}$  l'opérateur de perturbation.

Dans la théorie des perturbations de Møller-Plesset à l'ordre n (MPn), l'hamiltonien à l'ordre zéro est donné par la relation suivante :

$$\hat{H}^{(0)} = \sum_{i=1}^n \hat{f}_i \quad (\text{II.1.30})$$

où  $f_i$  est défini par la relation (II.1.15)

L'énergie du système vaut :

$$E^{MPn} = E^{(0)} + E^{(1)} + E^{(2)} + \dots + E^{(n)} \quad (\text{II.1.31})$$

$$\text{Avec } E_{HF} = E^{(0)} + E^{(1)} \quad (\text{II.32})$$

Donc la plus petite correction apportée à l'énergie HF est obtenue en limitant le développement à l'ordre 2, avec :

$$E^{(2)} = \sum_{i \neq 1} \frac{\langle \psi_0 | V | \psi_i \rangle \langle \psi_i | V | \psi_0 \rangle}{E_0 - E_i} \quad (\text{II.1.33})$$

La méthode MP2 est la plus utilisée de la série. Il a été montré qu'augmenter l'ordre perturbatif n'assure pas une amélioration des résultats même au prix d'un fort coût de calcul.

## II.2. La théorie de la fonctionnelle de la densité

### II.2.1. Théorie de la Fonctionnelle de la Densité indépendante du temps (DFT)

#### II.2.1.1. Généralités

Les méthodes post Hartree-Fock qui peuvent être d'une grande précision sont particulièrement coûteuses en temps de calcul et en mémoire. Leur utilisation est donc réservée à des systèmes de taille limitée. La DFT est une alternative à ces méthodes. Dans la DFT, développée selon les deux théorèmes de Hohenberg et Kohn, l'énergie électronique va être déterminée complètement par la densité électronique  $\rho$ . En effet, Hohenberg et Kohn, inspirés du modèle de Thomas-Fermi, ont montré que le potentiel externe, dans lequel se meuvent des particules en interaction, est déterminé par la densité électronique. De ce fait, l'énergie est une fonctionnelle de la densité :

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] \quad (\text{II.2.1})$$

où  $T(\rho)$  est l'énergie cinétique des électrons,  $V_{ee}[\rho]$  est l'énergie de répulsion coulombienne électron-électron et  $V_{Ne}[\rho]$  est l'énergie d'attraction coulombienne électron-noyau ayant comme expression :

$$V_{Ne}[\rho] = - \sum_a^N \int \frac{Z_a(R_a)\rho(r)}{|R_a-r|} dr \quad (\text{II.2.2})$$

$T[\rho] + V_{ee}[\rho]$  est une inconnue que l'on nomme fonctionnelle de Hohenberg-Kohn ( $F_{HK}[\rho]$ ).

La répulsion coulombienne électron-électron est séparée en deux termes :

$$V_{ee}[\rho] = J[\rho] + E_{ncl}[\rho] \quad (\text{II.2.3})$$

où  $J[\rho]$  est le terme coulombien ayant pour expression:

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \quad (\text{II.2.4})$$

et  $E_{ncl}[\rho]$  est un terme de contribution non classique comprenant la correction de self-interaction, l'échange et la corrélation électronique mais qui n'a pas d'écriture analytique connue.

Contrairement à la méthode Hartree-Fock, l'interaction coulombienne d'un électron avec lui-même peut ne pas se compenser exactement avec l'interaction d'échange. La DFT classique présente donc une erreur de self interaction. Celle-ci est le plus souvent négligeable, mais elle peut devenir conséquente pour des systèmes moléculaires spécifiques, dont certains ont été étudiés au cours de ces travaux de thèse (voir chapitre IV).

La fonctionnelle de Hohenberg et Kohn  $F_{HK}[\rho]$  a alors comme expression:

$$F_{HK}[\rho] = T[\rho] + J[\rho] + E_{ncl}[\rho] \quad (\text{II.2.5})$$

Le deuxième théorème de Hohenberg et Kohn énonce que la fonctionnelle  $E[\rho]$  donne l'énergie la plus basse seulement si la densité  $\rho$  est la même que celle de l'état fondamental. Ainsi, cette densité peut être déterminée à l'aide du principe variationnel:

$$E[\rho] \geq E_0[\rho_0] \quad (\text{II.2.6})$$

La limite du modèle de Hohenberg et Kohn est que la forme spécifique de  $F_{HK}[\rho]$  n'est pas connue. Toutefois, Kohn et Sham ont proposé une nouvelle expression de  $F_{HK}[\rho]$ , notée  $F[\rho]$ , en remplaçant le système réel, constitué de particules en interaction, par un gaz de particules sans interaction :

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \quad (\text{II.2.7})$$

où  $T_s[\rho]$  est l'énergie cinétique du système fictif constitué de particules sans interaction et  $E_{xc}[\rho]$  est l'énergie d'échange-corrélation qui a pour expression:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) \quad (\text{II.2.8})$$

La fonctionnelle  $E_{xc}[\rho]$ , dont l'expression analytique exacte est inconnue, est généralement décomposée en deux termes distincts, dont l'un est la fonctionnelle d'échange qui est associée

à l'interaction électron-électron de même spin  $E_X[\rho]$  et l'autre est la fonctionnelle de corrélation  $E_C[\rho]$  qui est associée à l'interaction électron-électron de spin opposé :

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \quad (\text{II.2.9})$$

Des écritures approchées de  $E_X[\rho]$  et  $E_C[\rho]$  doivent être introduites pour pouvoir calculer ces termes. Ces approximations faites pour estimer  $E_{XC}[\rho]$ , qui introduisent donc des erreurs à cette théorie exacte, dépendent de la classe des fonctionnelles considérées, dont les principes sont expliqués dans les parties suivantes de ce chapitre, et vont également être directement liées à la précision des fonctionnelles. Des articles et des revues testant la qualité des fonctionnelles ont été publiées, nous en citons quelques exemples.[7-18] Minimiser l'énergie revient à résoudre les équations de Kohn-Sham pour déterminer  $\Psi_i(r)$  appelée orbitale de Kohn-Sham et  $\varepsilon_i$  l'énergie associée à l'orbitale de Kohn-Sham. Le processus itératif permet la résolution de cet ensemble d'équations à partir des fonctions d'essai. La densité électronique  $\rho(r)$  du système fictif, imposée égale à celle du système réel, est exprimée par définition à partir des orbitales  $\Psi_i(r)$  :

$$\rho(r) = \sum_{i=1}^n |\Psi_i(r)|^2 \quad (\text{II.2.10})$$

### II.2.1.2. L'approximation de la densité locale (LDA)

Cette approche introduit l'idée d'un hypothétique gaz d'électrons uniforme dont la densité électronique  $\rho$  est traitée localement. Cette approximation est un bon modèle physique lorsque l'on étudie un atome ou une molécule où la densité électronique varie lentement tels que les métaux simples comme le sodium. La forme de la fonctionnelle d'énergie d'échange-corrélation est connue exactement:

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \varepsilon_{XC}(\rho(\vec{r})) d\vec{r} \quad (\text{II.2.11})$$

La partie échange a été définie par la relation de Bloch et Dirac :

$$\varepsilon_X(\rho(\vec{r})) = -\frac{3}{4} \sqrt{\frac{3\rho(\vec{r})}{\pi}} \quad (\text{II.2.12})$$

L'expression analytique de l'énergie de corrélation n'est pas connue. Toutefois, des simulations numériques de mécanique quantique Monte-Carlo de gaz d'électrons uniforme[19] permettent d'estimer la densité électronique donnant accès à l'énergie de corrélation. Plusieurs expressions de  $\varepsilon_c(\rho(\vec{r}))$  ont été proposées dans la littérature.[20-23]

En tenant compte de la polarisation de spin, la densité électronique a comme expression:

$$\rho(\vec{r}) = \rho_\alpha(\vec{r}) + \rho_\beta(\vec{r}) \quad (\text{II.2.13})$$

Dans les cas les plus simples, c'est à dire lorsque  $\rho_\alpha(\vec{r})$  est égale à  $\rho_\beta(\vec{r})$ , il y a compensation de spin. Mais dans les cas où  $\rho_\alpha(\vec{r})$  est différent de  $\rho_\beta(\vec{r})$ , il y a alors polarisation de spin. Cet effet est inclus dans l'énergie de corrélation en étant estimé par un paramètre de polarisation de spin:

$$\xi = \frac{\rho_\alpha(\vec{r}) - \rho_\beta(\vec{r})}{\rho(\vec{r})} \quad (\text{II.2.14})$$

On parlera alors de l'Approximation Locale de la Densité de Spin (LSDA). L'équation (2.11) devient alors :

$$E_{XC}^{LSDA}[\rho_\alpha, \rho_\beta] = \int \rho(\vec{r}) \varepsilon_{XC}(\rho_\alpha(\vec{r}) + \rho_\beta(\vec{r})) d\vec{r} \quad (\text{II.2.15})$$

Les fonctionnelles LDA ne traitent pas de la corrélation non-locale et en réalité le gaz d'électrons n'est pas homogène. Par conséquent, d'autres fonctionnelles ont été construites permettant de prendre en compte ces deux effets.

### II.2.1.3. L'approximation du gradient généralisé (GGA)

Dans cette génération de fonctionnelles, il n'est plus question de tenir compte seulement de la densité électronique  $\rho(\vec{r})$  en un point particulier de l'espace  $\vec{r}$  mais également de son gradient ( $\nabla\rho(\vec{r})$ ) afin de considérer l'inhomogénéité de la densité électronique. L'équation (2.15) devient alors :

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \quad (\text{II.2.16})$$



Toutefois, les fonctionnelles GGA ne sont pas des méthodes non-locales car elles dépendent de la densité et de ses dérivées premières en un point donné uniquement.

Les fonctionnelles d'échange GGA ont généralement pour premier terme les fonctionnelles d'échange LDA auquel s'ajoute un terme correctif :

$$E_X^{GGA} = E_X^{LDA} + \sum_{\sigma} \int F(s_{\sigma}) \rho_{\sigma}^{\frac{4}{3}}(\vec{r}) d\vec{r} \quad (\text{II.2.17})$$

avec  $\sigma$  correspondant au spin  $\alpha$  ou  $\beta$  et

$$s_{\sigma} = \frac{|\nabla \rho_{\sigma}(\vec{r})|}{\rho_{\sigma}^{\frac{4}{3}}(\vec{r})} \quad (\text{II.2.18})$$

étant le paramètre d'inhomogénéité local.

Les fonctionnelles d'échanges GGA sont couplées avec les fonctionnelles de corrélation, sachant que seules quelques combinaisons sont utilisées. Parmi les corrections possibles pour la fonctionnelle d'échange, on trouve parmi les plus utilisées celle de Becke[24] et celle de Perdew, Burke et Ernzerhof[25]. Concernant la fonctionnelle de corrélation, celle de Lee, Yang et Parr[26] et de Perdew sont les plus couramment utilisées à ma connaissance.[25,27,28]

#### II.2.1.4. Les fonctionnelles meta-GGA

Les fonctionnelles meta-GGA, qui sont une extension des fonctionnelles GGA, dépendent en plus explicitement du Laplacien de la densité  $\nabla^2 \rho_{\sigma}(\vec{r})$  ou de la densité de l'énergie cinétique  $\tau_{\sigma}(\vec{r})$ . [29-31] Il y a plusieurs définitions possibles de la densité d'énergie cinétique dont :

$$\tau_{\sigma}(\vec{r}) = \frac{1}{2} \sum_i |\nabla \psi_i(\vec{r})|^2 \quad (\text{II.2.19})$$

Par rapport aux GGA, les meta-GGA ont une dépendance semi-locale de la densité, de ses premières dérivées et des orbitales de Kohn-Sham en un point donné et dans un intervalle autour de ce point.

### II.2.1.5. Les fonctionnelles hybrides

Les fonctionnelles hybrides ont été construites afin de corriger l'erreur de self-interaction et de mieux traiter les phénomènes de corrélation non-locale due aux approximations faites dans les fonctionnelles dites classiques : LDA et GGA. En effet, par construction, ces deux classes de fonctionnelles ne peuvent pas prendre en compte les corrélations non locales tels que les interactions de type van der Waals et elles conduisent à une sur-délocalisation électronique. Pour considérer explicitement ce caractère non-local, Becke a introduit les fonctionnelles hybrides ; une partie de l'énergie d'échange exact du modèle Hartree-Fock est introduit dans les énergies d'échange et de corrélation de la DFT.[32] L'énergie d'échange corrélation a comme expression :

$$E_{XC} = E_{XC}^{DFT} + a_0(E_X^{exact} - E_X^{DFT}) \quad (\text{II.2.20})$$

Le paramètre  $a_0$  peut être paramétré théoriquement ou à l'aide de données expérimentales comme par exemple des énergies d'atomisation, des potentiels d'ionisation ou encore des affinités protoniques.

La fonctionnelle hybride la plus populaire est connue sous le nom de B3LYP et elle possède 20 % d'échange HF ;[26,32] cette fonctionnelle a été testée dans le chapitre IV et utilisée dans le chapitre V. La fonctionnelle la plus utilisée dans ce travail de thèse est M06[33] qui est une fonctionnelle meta-GGA hybride possédant 27 % d'échange HF. Elle a été hautement paramétrée pour reproduire les interactions non-covalentes.

### II.2.1.6. Les fonctionnelles corrigées à longue portée

L'ajout uniforme d'une partie de l'échange exacte du modèle Hartree-Fock permet de ne réduire que partiellement l'erreur de self-interaction. Ainsi, une autre catégorie de fonctionnelle a été développée, il s'agit des fonctionnelles corrigées à longue portée (Range-Separated Hybrid, RSH).[34-36] Le pourcentage d'échange Hartree-Fock, contrairement aux fonctionnels hybrides classiques, va augmenter en même temps que la distance inter-électronique. Cette croissance va être gouvernée par le paramètre d'atténuation, comme le montre l'expression de l'opérateur répulsion électronique:

$$\frac{1}{r_{12}} = \frac{\text{erf}(\omega r_{12})}{r_{12}} + \frac{1-\text{erf}(\omega r_{12})}{r_{12}} \quad (\text{II.2.21})$$

où  $\text{erf}$  est la fonction d'erreur de Gauss et  $r_{12} = |r_1 - r_2|$ . Le paramètre  $\omega$  définit la vitesse à laquelle la valeur de 1 sera atteinte par cette fonction d'erreur. Donc plus  $\omega$  sera grand, plus la transition de l'échange Kohn-Sham vers l'échange exact Hartree-Fock sera importante.

L'équation (II.2.21) a été généralisée par Yanai et *al.* en introduisant deux nouveaux paramètres  $\alpha$  et  $\beta$ [37]:

$$\frac{1}{r_{12}} = \frac{\alpha + \beta \text{erf}(\omega r_{12})}{r_{12}} + \frac{1 - [\alpha + \beta \text{erf}(\omega r_{12})]}{r_{12}} \quad (\text{II.2.22})$$

tel que  $0 \leq \alpha + \beta \leq 1$ ,  $0 \leq \alpha \leq 1$  et  $0 \leq \beta \leq 1$

Il existe également des fonctionnelles qui sont corrigées à moyenne portée;[38,39] leur opérateur répulsion électronique a pour expression :

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\omega_{SR} r_{12})}{r_{12}} + \frac{\text{erf}(\omega_{LR} r_{12})}{r_{12}} + \frac{[1 - \text{erf}(\omega_{LR} r_{12})] - [1 - \text{erf}(\omega_{SR} r_{12})]}{r_{12}} \quad (\text{II.2.23})$$

avec  $\omega_{SR}$  et  $\omega_{LR}$  les paramètres d'atténuation à courte et longue portée respectivement. Nous en avons testé une, il s'agit de HISSbPBE qui à 60% d'échange Hartree-Fock à « mi-portée » et 0 % d'échange HF à courte et longue portée. Les paramètres d'atténuation  $\omega_{SR}$  et  $\omega_{LR}$  valent 0.84 et 0.20 a.u respectivement.

### II.2.1.7. Les effets de dispersion

Il existe différentes approches pour prendre en compte les effets de dispersion qui sont manquantes dans la DFT. Parmi elles, (i) certaines fonctionnelles incluent explicitement un terme de dispersion non local, (ii) d'autres sont paramétrées, comme les fonctionnelles du Minnesota, et (iii) une autre approche consiste à ajouter un terme correctif empirique de dispersion. La discussion dans cette partie est centrée sur le point (iii), il s'agit d'une correction que nous avons utilisé dans le chapitre III, partie 2.

Grimme a proposé une forme de correction pour le traitement des effets de dispersion manquante dans la DFT et elle est actuellement la plus utilisée. Cette correction se fait par l'ajout de termes correctifs empiriques dépendant du nombre d'atomes  $N$ , du coefficient de dispersion pour la paire d'atome  $i$  et  $j$   $C_6^{ij}$ , de la distance inter-atomique  $R_{ij}$ , d'un facteur d'échelle  $s_6$  et d'une fonction d'amortissement  $f_{dmp}(R_{ij})$  tel que:

$$E_{DFT-D} = E_{KS-DFT} + E_D \quad (\text{II.2.24})$$

où  $E_D$  est l'énergie de dispersion et :

$$E_D = -s_6 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \quad (\text{II.2.25})$$

avec

$$f_{dmp}(R) = \frac{1}{1 + e^{-\alpha(\frac{R}{R_0}-1)}} \quad (\text{II.2.26})$$

$\alpha$  est un paramètre, et  $R_0$  est la somme des rayons atomiques de Van der Waals.[40]

Cette correction de dispersion a été affinée par Grimme en incluant un terme de non-additivité à trois corps noté  $E^{ijk}$  ;[41] la correction de dispersion résultante est qualifiée de correction de troisième génération « D3 »:

$$E_{disp} = E^{(2)} + E^{(3)} \quad (\text{II.2.27})$$

avec

$$E^{(2)} = -\sum_{ij} \sum_{n=6,8,10,\dots} s_n \frac{C_n^{ij}}{r_{ij}^n} f_{d,n}(R_{ij}) \quad (\text{II.2.28})$$

tel que

$$f_{d,n}(R_{ij}) = \frac{1}{1 + 6(\frac{R_{ij}}{s_{r,n}R_0})^{-a_n}} \quad (\text{II.2.29})$$

où  $a_n$  est un paramètre qui est ajusté. Le dernier terme de  $E_{disp}$  a pour expression :

$$E^{(3)} = \sum_{ijk} f_{d,(3)}(\overline{R_{ijk}}) E^{ijk} \quad (\text{II.2.30})$$

Plus récemment, Tkatchenko et Scheffler ont calculé d'une manière non-empirique les coefficients de dispersion  $C_6$ . [42]

Dans ce travail de thèse, la correction de dispersion de troisième génération a été utilisée avec la fonctionnelle B3LYP dans le chapitre III partie 2 (B3LYP-D3).

## ***II.2.2. Théorie de la Fonctionnelle de la Densité Dépendante du temps (TD-DFT)***

Lorsque l'on veut étudier les états excités d'une molécule en DFT, il n'est pas possible d'étendre simplement les équations de Kohn-Sham. Pour y remédier, il faut utiliser une autre approche qui est la théorie de la fonctionnelle de la densité dépendante du temps (TD-DFT).

Le théorème de Runge et Gross établit,[43] de façon similaire au premier théorème de Hohenberg et Kohn, qu'il existe une relation directe entre la densité électronique dépendante du temps et le potentiel externe  $v_{ext}$  auquel il est soumis. Dès lors, le potentiel et la fonction d'onde sont dépendants de la densité.

$$\rho(\vec{r}, t) \Leftrightarrow v_{ext}(\vec{r}, t) \Leftrightarrow \Psi(\vec{r}, t) \quad (\text{II.2.31})$$

Selon le premier théorème de Runge et Gross,[43] pour tout potentiel extérieur qui peut se développer en série de Taylor par rapport au temps autour de  $t=t_0$ , il existe une relation directe entre le potentiel  $v_{ext}(\vec{r}, t)$  et la densité électronique dépendante du temps  $\rho(\vec{r}, t)$  à une constante additive près. Une approche similaire à celle de Kohn-Sham consiste à choisir un système fictif d'électrons sans interaction, soumis à un potentiel effectif local dépendant du temps  $v_{KS}(\vec{r}, t)$  dont la forme de la densité  $\rho_s(\vec{r}, t)$  est identique à celle du système réel  $\rho(\vec{r}, t)$ . L'hamiltonien du système sans interaction a alors pour expression :

$$\hat{H}_s = - \sum_{i=1}^n \frac{\nabla_i^2}{2} + \sum_{i=1}^n v_{KS}(\vec{r}_i, t) \quad (\text{II.2.32})$$

et l'état associé est représenté par un déterminant de Slater qui résulte d'une combinaison linéaire d'orbitales mono-électroniques (II.34) qui sont solution des équations de Kohn-Sham dépendantes du temps (II.33) :

$$v_{KS}(r, t; \rho(\vec{r}, t))\varphi_i(\vec{r}, t) = [i \frac{\partial}{\partial t} + \frac{\nabla^2}{2}] \varphi_i(\vec{r}, t) \quad (\text{II.2.33})$$

La densité exacte dépendante du temps peut être calculée de la manière suivante:

$$\rho(\vec{r}, t) = \rho_s(\vec{r}, t) = \sum_{i=1}^n |\varphi_i(\vec{r}, t)|^2 \quad (\text{II.2.34})$$

et le potentiel effectif local peut-être décomposé de la façon suivante:

$$v_{KS}(\vec{r}, t) = v_{ext}(\vec{r}, t) + v_J(\vec{r}, t; \rho(\vec{r}, t)) + v_{XC}(\vec{r}, t; \rho(\vec{r}, t)) \quad (\text{II.2.35})$$

avec  $v_{ext}(r, t)$  le potentiel externe crée par les noyaux,  $v_J(r, t)$  le potentiel coulombien classique et  $v_{XC}(r, t)$  le potentiel d'échange corrélation non local. De là, on déduit l'énergie du système:

$$E[\rho](t) = T_s[\rho(\vec{r})] + \int v_{KS}[\rho(\vec{r}, t)]\rho(\vec{r}, t)dr \quad (\text{II.2.36})$$

Comme en DFT,  $v_{XC}(r, t)$  est inconnu et il est nécessaire d'introduire des approximations. La première est l'approximation adiabatique de la densité locale, qui considère que le noyau d'échange corrélation dépendant du temps peut être remplacé par un noyau statique sous l'hypothèse que le potentiel externe varie lentement dans le temps. L'objectif est de pouvoir décrire les états excités avec la même fonctionnelle que l'état fondamental. Dans ce cadre, si la perturbation extérieure ( $\delta v_{ext}$ ) qui agit sur la fonction d'onde de l'état fondamental est faible, alors la théorie de la réponse linéaire peut être appliquée.

En 1995, Casida a proposé une approximation de la théorie de la réponse linéaire pour déterminer de façon plus simple et moins coûteuse en temps de calcul les propriétés dynamique d'un système. Une première approche consiste à exprimer la réponse linéaire de la densité à partir de la fonction réponse  $\chi$  d'un système de particules en interaction à une perturbation. La méthode des perturbations permet d'obtenir une expression simplifiée de  $\chi$  sous la forme d'une somme sur tous les états excités considérés. Dans une seconde approche, la densité électronique est développée en série de Taylor par rapport au potentiel externe, mais seuls l'ordre 0 et le premier ordre sont considérés :

$$\rho(\vec{r}, t) = \rho_0(\vec{r}) + \delta\rho(\vec{r}, t) \quad (\text{II.2.37})$$

La densité  $\rho_0(\vec{r})$  correspond à celle du système non perturbé et est calculée en utilisant les équations de Kohn-Sham indépendantes du temps. La variation de la densité électronique  $\delta\rho(\vec{r}, t)$  a comme expression :

$$\delta\rho(\vec{r}, t) = \int \chi(\vec{r}, \vec{r}', t) \delta v_{ext}(\vec{r}', t) d\vec{r}' \quad (\text{II.2.38})$$

La fonction  $\chi$  n'est pas connue exactement mais elle est approximée par la réponse linéaire d'un système sans interactions  $\chi_s(r, r', t)$ . De même, un potentiel effectif  $\delta v_{eff}$  se substitue à la perturbation extérieure :

$$\delta\rho(\vec{r}, t) = \int \chi_s(\vec{r}, \vec{r}', t) \delta v_{eff}(\vec{r}', t) d\vec{r}' \quad (\text{II.2.39})$$

L'expression simplifiée de  $\chi$  sous la forme d'une somme sur tous les états excités est obtenue de la même manière que la première approche.

## II.3. Les bases d'orbitales atomiques

Comme nous l'avons vu précédemment, les orbitales atomiques et moléculaires résultent d'une combinaison linéaire de fonctions de base mono-électroniques, sachant qu'une base d'orbitales atomiques est définie pour chaque atome de la molécule.

Les fonctions de Slater sont les mieux adaptées pour reproduire les orbitales atomiques, en particulier au voisinage des noyaux. La base minimale correspond à l'utilisation d'une fonction de Slater pour décrire une orbitale atomique. Pour avoir une description plus précise des orbitales atomiques, il est possible de combiner plusieurs fonctions de Slater. Mais le produit de deux fonctions de Slater n'est plus une fonction de Slater, ce qui rend difficile le calcul des intégrales biélectronique. C'est pour cette raison que les fonctions Gaussiennes et leurs combinaisons sont le plus couramment utilisées en chimie quantique.

Les Gaussiennes contractées sont certes moins précises aux voisinages des noyaux que les fonctions de Slater, mais elles sont moins coûteuses en temps de calcul. Les bases de Pople se présentent sous cette forme : N-XYG avec N le nombre de gaussiennes primitives pour les orbitales atomiques de cœur ; XY le nombre de gaussiennes contractées pour décrire les orbitales de valence et (X, Y) qui indiquent le nombre de gaussiennes  $\chi_p(r, \theta, \varphi)$  utilisé pour construire les gaussiennes contractées. Pour mieux prendre en compte les interactions de type

van der Waals et/ou la déformation du nuage électronique, on ajoute des fonctions diffuse et /ou de polarisation respectivement sur les atomes lourds et/ou sur les atomes d'hydrogène. En effet, les fonctions de polarisation sont importantes en terme de liaison chimique car elles permettent aux orbitales de changer de forme. Dans la base de Pople, on ajoutera alors des fonctions d sur les atomes lourds et des fonctions p sur les atomes d'hydrogène. Les fonctions diffuses sont nécessaires pour les systèmes ayant de faibles interactions intermoléculaires, alors on ajoutera un + dans la base de Pople pour additionner des fonctions diffuses sur les atomes lourds et ++ pour ajouter à la fois des fonctions diffuses sur les atomes lourds et sur les atomes d'hydrogène.

Dans cette thèse, nous avons utilisé la base de Slater TZP (chapitre III partie 2), et les bases de Pople suivantes : 6-31G(d) (chapitre V partie 2), 6-31+G(d,p) (chapitre III), 6-311++G(d,p) (chapitre III), 6311++G(2d,p) (chapitre IV et V partie 1) et 6-311++G(2d,2p) (chapitre III partie 1).

## II.4. Les méthodes composites Gaussian-n

Les méthodes dites composites sont basées sur une série de calculs de chimie quantique qui, combinés selon un principe d'additivité, conduisent à des termes énergétiques proches de leurs valeurs expérimentales. Il existe actuellement plusieurs classes de méthodes composites : CBS, Gaussian (notées Gn), Wn, ... Le principe général consiste tout d'abord à optimiser la géométrie de l'espèce considérée et à évaluer les corrections thermiques à 0 et 298 K (ou autre température) à un niveau de théorie « moyen » (i.e. B3LYP/6-31+G(d,p), ...). Ensuite, des méthodes post-Hartree Fock (MP2, MP4), d'interaction de configuration quadratique ou de « coupled cluster » sont utilisées avec des bases d'orbitales atomiques de tailles moyennes. Les effets d'élargissement de bases sont introduits via des calculs MP2 ou MP4. Une procédure d'extrapolation linéaire ou non-linéaire des données obtenues avec une base finie à une base complète infinie est utilisée dans certaines méthodes comme G4[44,45] et CBS,[46] respectivement. Les méthodes Gn sont probablement les plus utilisées, les premières ont été élaborées en 1989 (G1,[47,48]) et elles ont été améliorées au fil des ans (G2,[49-51] G3[52-54] et G4[44,45]). Dans toutes les méthodes Gn, l'énergie totale du système est de la forme :

$$E_0 = E_{geom} + \Delta E_{base} + \Delta E_{corr} + (\Delta E_{SO}) + E_{HLC} + \Delta E_{ZPE} \quad (\text{II.4.1})$$



$E_{geom}$  est l'énergie du système avant l'ensemble des corrections apportées. Les effets de bases (diffuse, polarisation, taille) sont pris en compte et le terme correctif  $\Delta E_{base}$  est ajouté en conséquence. Les effets de corrélation électronique et le couplage spin-orbite sont pris en compte dans les termes correctifs  $\Delta E_{corr}$  et  $\Delta E_{SO}$  respectivement. Ce dernier terme correctif a été considéré seulement à partir de la série G3. Tous les effets résiduels qui n'auraient pas été pris en compte sont inclus dans un terme correctif empirique  $E_{HLC}$  dans lequel les constantes A et B sont paramétrées de manière à obtenir les plus petites déviations moyennes absolues par rapport à un ensemble d'énergies expérimentales (enthalpies de formation, énergies d'ionisation, affinités protoniques) :

$$E_{HLC} = -An_{\beta} - B(n_{\alpha} - n_{\beta}) \quad (\text{II.4.2})$$

où  $n_{\alpha}$  et  $n_{\beta}$  les nombres d'électron de valence  $\alpha$  et  $\beta$ .

À titre d'exemple, les différentes étapes de calculs utilisées dans la méthode G3B3, qui a été utilisée dans le chapitre V partie 2, sont résumées dans le tableau II.1. La géométrie du système est optimisée au niveau de théorie B3LYP/6-31G(d). L'énergie du point zéro est obtenue à partir des calculs de fréquences vibrationnelles à ce même niveau de théorie et est calibré avec un facteur d'échelle de 0.96. Il est à noter que lors des calculs avec la théorie de perturbation de Møller-Plesset au quatrième ordre, l'approximation des orbitales de cœur gelées, « frozen core approximation » (abréviation FC), est appliquée.

L'énergie totale du système en G3B3 est :

$$E_0 = E(\text{MP4}(\text{FC})/6 - 31\text{G}(\text{d})) + \Delta E(+) + \Delta E(\text{pol}) + E(\text{QCI}) + \Delta E(\text{Large}) + \Delta E_{SO} + E_{HLC} + \Delta E_{ZPE} \quad (\text{II.4.3})$$

$E(\text{MP4}(\text{FC})/6 - 31\text{G}(\text{d}))$  est l'énergie du système à la géométrie B3LYP/6-31G(d).

$\Delta E(+)$  est la correction due aux fonctions diffuses et a pour expression :

$$\Delta E(+) = E[\text{MP4}(\text{FC})/6 - 31 + \text{G}(\text{d})] - E[\text{MP4}(\text{FC})/6 - 31\text{G}(\text{d})] \quad (\text{II.4.4})$$

$\Delta E(\text{pol})$  est la correction due aux fonctions de polarisation :

$$\Delta E(\text{pol}) = E[\text{MP4}(\text{FC})/6 - 31\text{G}(2\text{df}, \text{p})] - E[\text{MP4}(\text{FC})/6 - 31\text{G}(\text{d})] \quad (\text{II.4.5})$$

$E(QCI)$  est une correction introduisant l'interaction de configuration quadratique telle que

$$E(QCI) = E(QCISD(T, FC)/6-31G(d)) - E(MP4/6-31G(d)) \quad (\text{II.4.6})$$

L'effet dû à l'utilisation de très grande base est corrigé par :

$$\Delta E(Large) = E(MP2(full)/G3Large) - E(MP2(FC)/6-31G(2df,p)) - E(MP2(FC)/6-31+G(d)) + E(MP2/6-31G(d)) \quad (\text{II.4.7})$$

Les paramètres A, B, C et D de la correction empirique sont :

$$A = 6.760 \text{ mhartrees}$$

$$B = 3.233 \text{ mhartrees}$$

$$C = 6.786 \text{ mhartrees}$$

$$D = 1.269 \text{ mhartrees}$$

$\Delta E_{ZPE}$  est la correction ZPE obtenue en B3LYP/6-31G(d).

Tableau II.1. Bilan des étapes de calculs en G3B3.

géométrie	B3LYP/6-31G(d)
ZPE	B3LYP/6-31G(d)
Scale factor	0,96
Simples points énergie	MP4(FC)/6-31G(d)
	MP4(FC)/6-31+G(d)
	MP4(FC)/6-31G(2df,p)
	QCISD(T)/6-31G(d)
	MP2(full)/G3Large
HLC	
Molécules	$-An_\beta - B(n_\alpha - n_\beta)$
Atomes	$-Cn_\beta - D(n_\alpha - n_\beta)$

## II.5. Les méthodes semi-empiriques

### II.5.1. Généralités

Le goulot d'étranglement (en termes de temps de calcul) des méthodes ab-initio est le calcul des intégrales à 3 et 4 centres. Les méthodes semi-empiriques réduisent grandement le nombre d'intégrales calculées. Elles reposent toutes sur un certain nombre d'approximations, et principalement :

- Une séparation cœur/valence (électrons de cœur gelés et seuls les électrons de valence sont traités)
- Les fonctions de bases sont des orbitales de Slater (STO)
- Les intégrales de recouvrement sont négligées  $S_{ij} = \delta_{ij}$
- Toutes les intégrales à 3 et 4 centres sont négligées, certaines intégrales à 1 ou 2 centres sont également négligées, les autres sont obtenues par paramétrisation (données expérimentales, spectrales, thermodynamiques, calculs ab-initio).

## ***II.5.2. Méthodes PM6 et PM6-DH2***

La méthode PM6 a été proposée par Stewart[55]. Elle se différencie des autres méthodes semi-empiriques car elle a adopté les paramètres d'interaction diatomique noyau-noyau de Voityuk's[56] et utilise l'approximation de Thiel[57] en ajoutant des orbitales d principalement aux éléments hypervalents du groupe principal de la classification périodique. De plus, le paramétrage pour plus de 70 éléments a été effectué. Toutefois, PM6 manque de fiabilité dans la description des interactions non-covalentes, notamment pour l'énergie de dispersion et celle des liaisons hydrogène. Pour remédier à cela, des corrections ont été introduites. La correction de dispersion est similaire à celle proposée par Grimme en DFT :

$$\sum E_{disp} = - \sum f_{dmp} (r_{ij}, R_{ij}^0) \frac{C_6^{ij}}{r_{ij}^6} \quad (\text{II.5.1})$$

Les paramètres atomiques  $C_6^{ij}$  et  $R_{ij}^0$  correspondent au coefficient de dispersion pour la paire d'atomes ij et au rayon atomique de van der Waals respectivement. En incluant cette correction dans PM6, cette méthode sera notée PM6-D.

Plusieurs corrections existent pour les liaisons hydrogène.[58-60] La méthode de deuxième génération « H2 » consiste en la détermination des paramètres de liaisons hydrogène pour différents types d'atomes (donneurs ou accepteurs) dans diverses conformations. Au final, nous avons utilisé la méthode PM6-DH2 qui prend en compte les deux types de correction.

Pour cette méthode, il y a une erreur dans le gradient de l'énergie en ce qui concerne la géométrie. Il est donc préconisé d'optimiser les géométries avec PM6 et de faire des calculs d'énergie avec PM6-DH2. C'est cette stratégie que nous avons utilisée.

Récemment, une nouvelle version a été développée, il s'agit de PM7 qui possède de nouvelles paramétrisations qui donnent des résultats plus précis pour l'étude des éléments du groupe principal et des métaux de transition.[61] De nouvelles fonctions de dispersion et de nouveaux termes pour les liaisons hydrogène y sont intégrés.

## II.6. Exploration des surfaces d'énergie potentielle

### II.6.1. Exploration en mécanique classique

#### II.6.1.1. Les outils

##### II.6.1.1.1. Généralités

La mécanique quantique est très couteuse en temps de calcul, ce qui la restreint à une utilisation pour des systèmes de petite taille. Pour explorer des surfaces d'énergie potentielle de systèmes flexibles et complexes, il faut donc avoir recours à des méthodes plus simples. Les champs de forces, développés sur la base de la mécanique classique, permettent l'étude de systèmes de très grandes tailles comme par exemple les protéines, car l'énergie est calculée plus simplement qu'en mécanique quantique. Le terme « champ de forces » désigne la forme mathématique de l'énergie potentielle et les paramètres associés. Par exemple, pour des champs de forces usuels tels que AMBER[62] (*Assisted Model Building and Energy Refinement*) et CHARMM[63,64] (*Chemistry at HARvard Molecular Mechanics*), qui sont régulièrement utilisés pour l'étude de biomolécules, les atomes sont généralement considérés comme des charges ponctuelles reliés par des ressorts, et l'énergie potentielle a pour expression :

$$E_{MM} = E_{str} + E_{bnd} + E_{tor} + E_{vdW} + E_{ele}^{perm} \quad (II.6.1)$$

Dans cette expression, la somme  $E_{str} + E_{bnd} + E_{tor}$  représente les interactions entre atomes liés. Ces termes représentent la déformation des liaisons, des angles et des angles dièdres entre deux, trois et quatre atomes voisins respectivement et ont le plus souvent pour expression:

$$E_{str} = \sum_{liaisons} \frac{1}{2} k_b (r - r_0)^2 \quad (\text{II.6.2})$$

$$E_{bnd} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (\text{II.6.3})$$

avec  $k_b$  et  $k_\theta$  les constantes de force de l'oscillateur harmonique et,  $(r - r_0)$  et  $(\theta - \theta_0)$  les déviations par rapport aux positions d'équilibre de liaison  $r_0$  et d'angle  $\theta_0$  respectivement.

$$E_{tor} = \sum_{diedre} A [1 + \cos(n\tau - \varphi)] \quad (\text{II.6.4})$$

avec  $\tau$  l'angle de torsion,  $\varphi$  le facteur de phase,  $n$  le facteur de périodicité et  $A$  l'amplitude.

La somme  $E_{vdW} + E_{ele}^{perm}$  représente les interactions entre atomes non liés. Le terme de van der Waals  $E_{vdW}$  est souvent décrit par le potentiel de Lennard-Jones avec un terme attractif en  $\frac{1}{r_{ij}^6}$  et un terme répulsif en  $\frac{1}{r_{ij}^{12}}$ :

$$E_{vdW} = \sum_{i>j} -\frac{C_{ij}}{r_{ij}^6} + \frac{D_{ij}}{r_{ij}^{12}} \quad (\text{II.6.5})$$

Les interactions électrostatiques sont déterminées par l'interaction coulombienne entre des particules chargées  $q_i$  et  $q_j$  dans un milieu de permittivité relative  $\epsilon_r$ :

$$E_{ele}^{perm} = \sum_{i>j} -\frac{q_i q_j}{\epsilon_r r_{ij}} \quad (\text{II.6.6})$$

Les interactions à n-corps, telle que la polarisation, ne sont donc pas explicitement prises en compte. Les paramètres des différents termes énergétiques sont obtenus à l'aide de données expérimentales ou de calculs ab initio. Les champs de forces classiques sont semblables dans l'expression de  $E_{MM}$  mais diffèrent principalement par leur paramétrisation ou par l'ajout de terme énergétique voire de couplage. Pour CHARMM par exemple, le couplage Urey-Bradley est ajouté et correspond au couplage entre les vibrations des angles et des liaisons. Pour les champs de forces MMx, adaptés pour les molécules organiques et développés par Allinger et

al.,[65-68] des termes d'anharmonicité et de couplage entre les termes liés sont ajoutés à l'énergie potentielle.

Le champ de forces UFF (*Universal Force Field*) a été utilisé lors de nos calculs QM/MM. Il est adapté aux molécules organométalliques car il a l'avantage de posséder des paramètres pour l'ensemble des éléments de la classification périodique des éléments. Sa paramétrisation est générale grâce à de simples règles de combinaisons de paramètres atomiques que nous ne développerons pas ici. Son utilisation est donc simple mais sa précision est limitée.[69]

Les champs de forces classiques ont une précision suffisante lorsque la polarisation résultant de l'approche d'une autre molécule n'est pas l'interaction prépondérante. Dans le cas contraire, l'utilisation d'un champ de forces dit « polarisable » devient nécessaire. Au vu des systèmes étudiés, nous avons opté pour le champ de forces polarisable AMOEBA, développé par Ponder et ses collaborateurs.[70,71] De plus, AMOEBA est utilisé et développé au laboratoire LCM pour des molécules en phase gazeuse.

#### **II.6.1.1.2. AMOEBA, Atomic Multipole Optimized Energetics for Biomolecular Applications.**

Comme nous le verrons par la suite, AMOEBA bénéficie d'un traitement efficace des effets électrostatiques et de la polarisation. De plus, la transférabilité de ses paramètres, obtenus par calculs de chimie quantique, est un atout.[70] Ce champ de force est spécialement développé pour l'étude de molécules biologiques.[70-72] Jensen et ses collaborateurs ont montré la limitation des champs de force usuels dans l'étude d'acide aminés neutre N-acétylé tandis qu'AMOEBA a donné à la fois des structures fiables et une bonne reproduction de leurs énergies relatives.[73] Sa pertinence a été confirmée lors de l'étude de systèmes chargés comme la méthionine protonée et déprotonée[74], la gramicidine S[75] et les polyglycines et les polyalanines sodées.[76-78]

Tout comme les champs de forces classiques, l'énergie potentielle est la somme d'un potentiel pour les interactions liées et les interactions non liées. L'effet de polarisation est pris en compte dans le terme d'interaction électrostatique.

$$E_{\text{AMOEBA}} = E_{\text{str}} + E_{\text{bnd}} + E_{\text{str-bnd}} + E_{\text{oop}} + E_{\text{tor}} + E_{\text{vdW}} + E_{\text{ele}}^{\text{perm}} + E_{\text{ele}}^{\text{ind}} \quad (\text{II.6.7})$$

## Les interactions liées

Parmi les interactions liées, le formalisme de MM3[67] pour la déformation des liaisons, des angles et de leur couplage est conservé.

## Les interactions électrostatiques

Dans AMOEBA, les dipôles et les quadrupôles permanents sont utilisés en plus de l'interaction de deux charges ponctuelles. {charge, dipôle, quadrupôle} forme un multipôle atomique permanent. Pour un système donné, les multipôles atomiques sont représentés par une matrice  $M_i$  tel que :

$$M_i = [q_i, \mu_{i,x}, \mu_{i,y}, \mu_{i,z}, Q_{i,xx}, Q_{i,xy}, Q_{i,xz}, \dots, Q_{i,zz}]^t \quad (\text{II.6.8})$$

et l'énergie d'interaction électrostatique entre les atomes  $i$  et  $j$  séparés par une distance  $r_{ij}$  est:

$$E_{ele}^{perm}(r_{ij}) = M_i^t T_{ij} M_j \quad (\text{II.6.9})$$

où  $T_{ij}$  est la matrice d'interaction entre les multipôles des atomes  $i$  et  $j$ .

Les multipôles atomiques sont extraits à partir de calculs ab initio réalisés sur de petits systèmes en utilisant la méthode DMA (Distributed Multipole Analysis) de Stone.[79,80]

Pour chaque atome  $i$ , le dipôle induit formé a pour expression :

$$\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha} \quad (\text{II.6.10})$$

où  $E_{i,\alpha}$  représente le champ électrostatique sur un atome  $i$  et  $\alpha_i$  est la polarisabilité dipolaire atomique.  $E_{i,\alpha}$  est la somme de tous les champs électriques générés par les multipôles permanents et par les dipôles induits des autres atomes  $i$ . Cette équation devient :

$$\mu_{i,\alpha}^{ind} = \alpha_i (\sum_{\{j\}} T_{ij}^\alpha M_j + \sum_{\{j'\}} T_{ij'}^{\alpha\beta} \mu_{j',\beta}^{ind}) \quad (\text{II.6.11})$$

Cette équation peut être résolue de manière itérative pour tous les sites atomiques. Il est à noter que, à courte distance, l'interaction de polarisation est corrigée par un facteur d'amortissement noté  $\rho$  afin d'éviter une catastrophe de polarisation.[81]

$$\rho = \frac{3a}{4\pi} e^{-au^3} \quad (\text{II.6.12})$$

où

$$u = \frac{r_{ij}}{(\alpha_i \alpha_j)^{\frac{1}{6}}} \quad (\text{II.6.13})$$

est la distance effective et  $a$  est un paramètre sans dimension qui détermine la force d'amortissement.

### Les interactions de van der Waals

L'énergie pour les interactions de type van der Waals est donnée par l'équation suivante :

$$E_{vdW}(ij) = \varepsilon_{ij} \left( \frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left( \frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right) \quad (\text{II.6.14})$$

avec

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}})^2} \quad (\text{II.6.15})$$

où  $\varepsilon_{ij}$  est la profondeur du puits de potentiel pour les atomes  $i$  et  $j$ . De plus :

$$\rho_{ij} = \frac{R_{ij}}{R_{ij}^0} \quad (\text{II.6.16})$$

où  $R_{ij}^0$  est la distance du minimum d'énergie entre les atomes  $i$  et  $j$  dépendant de la distance du minimum d'énergie pour les paires homoatomiques  $ii$  et  $jj$  selon la relation :

$$R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2} \quad (\text{II.6.17})$$

Les paramètres de van der Waals sont ajustés en utilisant les données expérimentales obtenues en phase gazeuse et en phase condensée.

### Classification des atomes

Pour définir l'ensemble des paramètres AMOEBA, chaque atome est classé selon une classe et un type d'atome. Les paramètres contenus dans la classe d'atome permettent de définir les interactions liées ainsi que celles de van der Waals. Pour une même classe, il existe différents



type d'atomes qui permettent de définir les interactions électrostatiques et de polarisation et de s'adapter à l'environnement chimique. Des protocoles d'optimisation de paramètres AMOEBA sont régulièrement proposés dans la littérature.[70,72,82,83]

Afin d'explorer de manière exhaustive la surface d'énergie potentielle de nos systèmes, nous avons utilisé ce champ de force avec différents outils pour générer les différentes conformations adoptées par le système : dynamique moléculaire, simulations Monte Carlo et saut de bassins. En fonction des systèmes, ces méthodes ont été utilisées de façon complémentaire afin d'identifier le plus possible de minima de basse énergie sur la surface d'énergie potentielle.

### ***II.6.1.1.3. Dynamique moléculaire classique***

Pour un système composé de N atomes, la dynamique du système étudié est gouvernée par les équations de Newton afin de suivre l'évolution des coordonnées de chaque atome dans le temps :

$$\sum \vec{F}_{ext} = m \frac{d^2 r}{dt^2} = m \vec{a} \quad (\text{II.6.18})$$

Ainsi, on a un système de N équations couplées à résoudre. Chaque force exercée par chaque atome sur un atome en particulier peut être calculée grâce à un potentiel effectif classique qui est l'énergie potentielle AMOEBA dans nos travaux. Afin de résoudre les N équations couplées, les positions des noyaux initiales sont définies par la géométrie initiale du système et les vitesses sont générées de manière aléatoire. Pour calculer les positions des atomes aux instants  $t$  ( $r_i$ ) et  $t+\Delta t$  ( $r_{i+1}$ ), le temps est discrétisé et un algorithme numérique est utilisé pour résoudre les équations du mouvement. Les trois algorithmes les plus utilisés sont les algorithmes de Verlet, Verlet-vitesse et de Beeman, ce dernier étant utilisé dans ce travail de thèse.

Les positions et les vitesses calculées simultanément par l'algorithme de Beeman sont obtenues par les équations ci-dessous.[Beeman1976]

$$r_{i+1} = r_i + v_i \Delta t + \frac{1}{6} [4a_i - a_{i-1}] \Delta t^2 \quad (\text{II.6.19})$$

$$v_{i+1} = v_i + \frac{1}{6} [2a_{i+1} + 5a_i - a_{i-1}] \Delta t \quad (\text{II.6.20})$$

Cet algorithme est l'algorithme par défaut pour la dynamique moléculaire dans le logiciel TINKER.

#### ***II.6.1.1.4. Simulation Monte Carlo***

La méthode Monte Carlo (MC) est une méthode stochastique qui permet de générer aléatoirement des configurations représentatives du système selon une distribution de Boltzmann  $e^{-\frac{\Delta E}{kT}}$ . Deux configurations successives sont liées entre elles par des mouvements arbitraires. Cela permet d'échantillonner au mieux la surface d'énergie potentielle du système étudié. La simulation commence par un ensemble de coordonnées raisonnable pour l'ensemble des noyaux ( $C_i$ ). La géométrie suivante ( $C_{i+1}$ ) est incluse dans l'ensemble des configurations uniquement si son énergie est plus basse que celle de la conformation précédente. Si ce n'est pas le cas, une distribution de Boltzmann est calculée et est comparée à un nombre aléatoire compris entre 0 et 1. Si la valeur obtenue est plus petite que le nombre choisi arbitrairement alors la configuration est conservée, sinon elle est refusée. Si elle est acceptée, la configuration ( $C_i$ ) est ajoutée à l'ensemble des configurations obtenues et de nouveaux mouvements arbitraires sont effectués.

Dans ce travail, nous avons utilisé les simulations Monte Carlo dans le chapitre III partie 1 afin d'essayer de garantir que la structure de plus basse énergie a bien été identifiée par la dynamique moléculaire et/ou le saut de bassins (voir partie 6.1.1.5). C'était un moyen également de vérifier que pour une gamme d'énergie donnée, nous ne trouvions pas de nouveaux minima en MC.

#### ***II.6.1.1.5. Saut de bassins***

Le saut de bassins est une méthode qui transforme la surface d'énergie potentielle en « escalier », le minimum global et les énergies relatives des autres minima restant inchangés.[84] Seules les barrières sont supprimées de la surface. Chaque plateau contient un ensemble de configurations qui sont reliées au même minimum après optimisation de géométrie. L'énergie transformée d'une configuration  $C_i$  est donc défini par :

$$\tilde{E}(C_i) = \min\{E(C_i)\} \quad (\text{II.6.21})$$

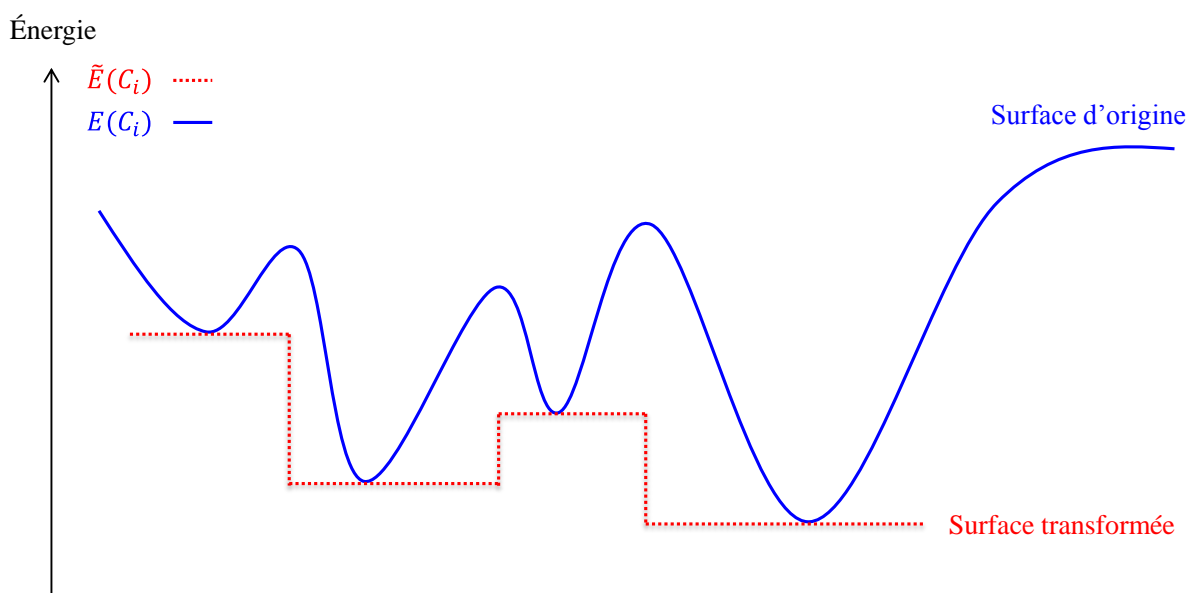


Figure II.2. Schéma à une dimension illustrant la transformation de la surface d'énergie potentielle.

L'exploration de la surface en forme « d'escalier » est aisée et peut être faite en utilisant des méthodes stochastiques (Minimisation Monte Carlo ou bien des algorithmes génétiques). Dans TINKER, la recherche des structures est effectuée par des mouvements de torsions qui correspondent à des modes de vibration et dont le nombre est à définir.

### II.6.1.2. Stratégies d'exploration de surface pour les systèmes micro-hydratés

Afin d'explorer de manière exhaustive la surface d'énergie potentielle de système biologique micro-hydraté, deux stratégies ont été largement utilisées dans la littérature. La première est basée sur la construction d'un arbre type « Darwin » et la seconde est basée sur des processus successifs d'optimisation de plus en plus performants. Une description sur le fonctionnement de ces deux approches est donnée ci-dessous.

#### II.6.1.2.1. L'approche arbre de Darwin

La première étape consiste à établir un arbre type Darwin en identifiant les différents sites d'hydratation des molécules de  $H_2O$  pour une conformation donnée appartenant à la génération  $G_0$  (ensemble des structures non hydratées). Pour les molécules flexibles, l'identification des structures de  $G_0$  relève donc d'un challenge. De  $G_0$  découle les générations suivantes, notée  $G_n$ , solvatées par  $n$  molécules de  $H_2O$  et obtenues après l'ajout d'une

molécule sur  $G_{n-1}$  pour les différents sites d'hydratation possibles. Michaux et al.[85] ont développé cette stratégie pour étudier le système  $\text{GlyH}^+\cdots\text{H}_2\text{O}$ . Depuis, d'autres études sur la micro-solvatation de petits acides aminés protonés ( $\text{GlyH}^+(\text{H}_2\text{O})_{n=0,1,2,3,4}$ ,[86]  $\text{AlaH}^+(\text{H}_2\text{O})_{n=0,1,2,3}$ ,[87]  $\text{ProH}^+(\text{H}_2\text{O})_{n=0,1,2,3}$ [88]) ont été basées sur cette stratégie. De ces études est ressorti que la conformation la plus stable de  $G_{n-1}$  se retrouve être la plus stable dans la génération  $G_n$ . Ces observations ont été confirmées lors de l'étude d'acides aminés déprotonés flexibles ou non,  $\text{Val}\cdots\text{H}_2\text{O}$ ,  $\text{Ala}\cdots\text{H}_2\text{O}$ ,  $\text{Asp}\cdots\text{H}_2\text{O}$  et  $\text{Gln}\cdots\text{H}_2\text{O}$ ,[89] et sur le phénol complexé à la fois par  $n$   $\text{H}_2\text{O}$ ,  $n=1,2,3$  et 4, et par un cation métallique ( $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Li}^+$  et  $\text{Mg}^{2+}$ ).[90] Mais cette stratégie doit être utilisée avec la plus grande précaution car comme nous l'avons vu dans le chapitre 1, l'hydratation peut induire des changements de conformation (étendue ou repliée) et/ou de site de protonation.

#### II.6.1.2.2. L'approche hiérarchique

Dans la majorité des cas, les géométries de départ sont générées par dynamique moléculaire, Monte Carlo et/ou saut de bassins, en utilisant un champ de forces adapté qui décrit bien le système en terme d'énergie et de géométrie. Les structures de plus basses énergies sont sélectionnées puis ré-optimisées en utilisant des outils de mécanique quantique. Deux scénarii sont envisageables : le premier consiste en l'utilisation d'un haut niveau de théorie directement,[91,92] le second inclut d'abord un ou plusieurs niveaux intermédiaires afin d'éliminer le plus grand nombre de structures. Les niveaux intermédiaires peuvent être, soit un bas niveau de calcul *ab initio*, comme la méthode Hartree-Fock,[93] ou bien une méthode semi-empirique[94] ou bien une DFT avec une petite base, auquel cas le(s) niveau(x) suivant(s) sera(ont) la même DFT mais en augmentant la taille de la base,[95-97] ou bien un plus haut niveau de théorie.

Dans une récente étude concernant la micro-solvatation de la glycine protonée, les approches hiérarchique et arbre de type Darwin ont été réalisées en parallèle.[91] Elles ne permettent pas individuellement d'identifier toutes les structures de plus basses énergies, mais ensemble elles sont complémentaires.

## ***II.6.2. Les Chemins Réactionnels***

Pour pouvoir expliquer la formation de molécules, l'élucidation du mécanisme réactionnel devient essentielle. Les mécanismes étudiés au cours de cette thèse sont principalement des migrations de proton ou de radical H. Ainsi, à partir des réactifs de départs, il est plutôt aisé d'identifier la nature de l'état de transition en variant la distance  $X-H\cdots Y$  tel que  $X\cdots H\cdots Y$ . Pour un état de transition, le mode normal de vibration est la coordonnée de vibration. Ainsi, nous avons suivi à partir de chaque état de transition un chemin de réaction nommé Coordonnée de Réaction Intrinsèque (IRC), afin de nous assurer que l'état de transition identifié est bien connecté aux réactifs et aux produits. En effet, l'IRC permet de suivre le chemin de plus grande pente dans deux directions différentes, l'une vers les réactifs et l'autre vers les produits. Tous les IRC ont été réalisés avec le logiciel Gaussian09 qui utilise un nouvel algorithme HPC (« Hessian-based-Predictor-Corrector »).[98-100]

## ***II.6.3. Remarques***

Les approches hiérarchique (en utilisant les simulations Monté Carlo, de dynamique moléculaire et de sauts de bassins associées au champ de forces AMOEBA) et arbre de Darwin ont été utilisées afin d'explorer le plus exhaustivement possible la surface d'énergie potentielle des molécules (acides aminés acétylés protonés, aminoalcool neutre et protoné) microhydratées, la stratégie utilisée étant expliquée en détail dans le chapitre III partie 1. Compte tenu de la faible flexibilité de la glycine bêtaïne seule et de la grande flexibilité des clusters de glycine bêtaïne (chapitre III partie 2), les structures ont été générées uniquement par dynamique moléculaire.

Des IRC ont été réalisés pour tous les états de transition obtenus dans ce travail de thèse.

## II.7. Les méthodes hybrides QM/MM

### II.7.1. Généralités

Comme nous l'avons vu précédemment, la mécanique classique est un outil pour traiter des systèmes étendus. Toutefois, elle ne permet pas de traiter explicitement tous les phénomènes électroniques. De plus, l'étude de la réactivité au sein de systèmes biologiques de grande taille n'est pas possible en mécanique classique mais aussi en mécanique quantique dû à la limitation de la taille du système. Pour remédier à ce problème, en 1976, Warshel et Levitt ont proposé une approche « hybride » pour l'étude de réactions enzymatiques (enzyme-substrat) en traitant le clivage de la liaison et la redistribution de charge du substrat en mécanique quantique, alors que les effets stériques et électrostatiques étaient traités en mécanique classique.[101] Depuis, des recherches ont été menées activement dans le but de développer différentes approches hybrides. Dans le cadre de la réactivité, le site réactif est généralement traité par la mécanique quantique et le reste du système par la mécanique classique.[102-106] Cette méthode hybride est appelée Mécanique Quantique-Mécanique Moléculaire que l'on notera QM/MM. Le développement de ces méthodes a permis à Martin Karplus, Michael Levitt et Arieh Warshel d'obtenir le prix Nobel 2013 de Chimie.[107]

Puisque le système est scindé en deux régions,  $R_{MM}$  et  $R_{QM}$ , l'énergie totale du système inclura en plus des énergies obtenues en QM et en MM, un terme énergétique correspondant à l'interaction entre les deux régions notée  $E^{QM/MM}$  :

$$E_{tot} = E^{QM} + E^{MM} + E^{QM/MM} \quad (\text{II.7.1})$$

$E^{QM}$  se calcule avec les méthodes *ab initio*, DFT et aussi les méthodes semi-empiriques.  $E^{MM}$  se calcule avec les champs de forces usuels.

Il existe plusieurs méthodes pour calculer l'énergie d'interaction entre les deux régions ainsi que pour décrire la frontière. Dans nos travaux, nous avons utilisé l'approche ONIOM.

### II.7.2. L'approche ONIOM

L'approche ONIOM a été utilisée en scindant en deux parties le système étudié. Les liaisons frontières sont définies suivant l'approche « link-atom » qui consiste à remplacer l'atome frontière de l'autre région par un atome d'hydrogène. Le système complet sera dit système

« réel », alors que le site actif sera qualifié de système « modèle ». Le système modèle est traité à un « haut » niveau de théorie (QM) et le système réel est traité à un « bas » niveau de théorie (MM). L'énergie totale du système est :

$$E_{tot,ONIOM} = E_{réel}^{bas} + E_{modèle}^{haut} - E_{modèle}^{bas} \quad (II.7.2)$$

## II.8. Les effets de solvation

Pour considérer le solvant, l'idéal serait de le traiter de façon explicite avec un très grand nombre de molécules d'eau, mais ce qui est inenvisageable en chimie quantique. De ce fait, les effets du solvant peuvent être considérés par (i) quelques molécules de H<sub>2</sub>O explicites dont on cherche à déterminer localement les propriétés thermochimiques, structurales ou mécanistique (chapitre III partie 1 et chapitre V partie 2), (ii) un continuum (chapitre V partie 2) ou (iii) une combinaison des deux points précédents en QM (chapitre V partie 2) ou QM/MM.

Dans le chapitre V partie 2, nous avons utilisé le modèle « Conductor-like Polarizable Continuum Model » (C-PCM).[108,109] Le système est placé dans une cavité vide mais l'extérieur est un milieu continu polarisable possédant certaines propriétés macroscopiques du solvant, comme par exemple sa constante diélectrique. La cavité est définie par des sphères de van der Waals centrées sur chaque atome et la polarisation du solvant est exprimée en terme de densité de charge placée à la surface de la cavité. Cette polarisation induit un champ électrique dans la cavité et agit donc sur le système. La forme de la cavité est plus réaliste puisqu'elle est divisée en tesserae. Le terme « conductor-like » est dû au fait que chaque tesserae inclut à son centre des charges ponctuelles polarisables. Il existe plusieurs types de cavités (UFF, UA0, UAHS, UAKS, Pauling, Bondi) dont les rayons des sphères de van der Waals sont optimisés avec un niveau de calcul. Dans notre étude, nous avons utilisé les cavités (i) UFF, dans laquelle les rayons sont ceux obtenus avec le champ de forces UFF et les atomes d'hydrogène sont des sphères individuelles ; et (ii) UAKS dans laquelle les rayons sont optimisés avec la méthode PBE0/6-31G(d) (voir Chapitre V partie 2 pour le choix des cavités).

L'énergie libre de solvation a pour expression :

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{disp-rep} + \Delta G_{elec} \quad (II.8.1)$$

$\Delta G_{elec}$  représente les interactions électrostatiques entre le système et le solvant incluant la polarisation.  $\Delta G_{disp-rep}$  prend en compte les effets de dispersion et/ou de répulsion et  $\Delta G_{cav}$  est la contribution énergétique nécessaire pour que la cavité se forme dans le solvant, ce qui est entropiquement défavorable.

## II.9. Outils d'analyse

### II.9.1. Les grandeurs thermodynamiques

Les grandeurs thermodynamiques que nous avons utilisées tout au long de ce travail sont : l'énergie potentielle ( $E_0$ ), l'énergie du point zéro (ZPE), l'entropie ( $S^\circ(T)$ ), l'enthalpie ( $H^\circ(T)$ ) et l'enthalpie libre ( $G^\circ(T)$ ). Ces grandeurs se définissent à partir de la fonction de partition du système  $q(V,T)$ . Cette dernière peut s'écrire comme le produit des fonctions de partition électronique, translationnelle, vibrationnelle et rotationnelle du système:

$$q(V,T) = q_e q_{trans} q_{vib} q_{rot} \quad (\text{II.9.1})$$

La fonction de partition électronique a pour expression :

$$q_e = \sum_i \omega_i e^{-\frac{\varepsilon_i}{k_B T}} \quad (\text{II.9.2})$$

où  $\omega_i$  est la dégénérescence du niveau d'énergie  $i$  et  $\varepsilon_i$  est son énergie. Les états électroniques excités sont supposés être inaccessibles car leur niveau d'énergie  $\varepsilon_i$  par rapport à l'état fondamental est nettement supérieur à  $k_B T$ , ainsi  $\sum_{i>0} \omega_i e^{-\frac{\varepsilon_i}{k_B T}} \rightarrow 0$  et donc :

$$q_e = \omega_0 \quad (\text{II.9.3})$$

La fonction de partition translationnelle est donnée par la relation suivante :

$$q_{trans} = \left( \frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} V \quad (\text{II.9.4})$$

avec



$$V = \frac{k_B T}{P} \quad (\text{II.9.5})$$

dans l'hypothèse d'un gaz parfait et m est la masse du système étudiée.

En considérant le point de référence zéro, la fonction de partition vibrationnelle est donnée par :

$$q_{vib} = \prod_k \frac{e^{-\frac{\theta_{v,k}}{2T}}}{1 - e^{-\frac{\theta_{v,k}}{T}}} \quad (\text{II.9.6})$$

avec

$$\theta_{v,k} = \frac{h\nu_k}{k_B} \quad (\text{II.9.7})$$

la température vibrationnelle de chaque mode (si N désigne le nombre d'atomes de la molécule, k = 3N-6 modes vibrationnelles pour une molécule non-linéaire et 3N-5 pour une molécule linéaire). L'énergie du point zéro (ZPE) vaut :  $\frac{\theta_{v,k}}{2}$ .

Les fonctions de partitions rotationnelles d'une molécule linéaire et non-linéaire sont respectivement:

$$q_{rot} = \frac{1}{\sigma_r} \left( \frac{T}{\theta_r} \right) \quad (\text{II.9.8})$$

avec

$$\theta_r = \frac{h^2}{8\pi^2 I k_B} \quad (\text{II.9.9})$$

où I est le moment d'inertie

et

$$q_{rot} = \frac{1}{\sigma_r} \left( \frac{T^{\frac{3}{2}}}{(\theta_{r,x} \theta_{r,y} \theta_{r,z})^{\frac{1}{2}}} \right) \quad (\text{II.9.10})$$

dans l'approximation des particules indépendantes (gaz parfait).

L'énergie interne thermique du système a pour expression :

$$E_{int} = N k_B T^2 \left( \frac{\partial \ln q(V,T)}{\partial T} \right)_V \quad (\text{II.9.11})$$

et celle de l'entropie est :

$$S^\circ(T) = R[\ln(q(V,T)e) + T(\frac{\partial \ln q(V,T)}{\partial T})_V] \quad (\text{II.9.12})$$

De ces deux formules, on peut en déduire l'enthalpie et l'enthalpie libre de Gibbs :

$$H^\circ(T) = E_0 + E_{int} + k_B T \quad (\text{II.9.13})$$

et

$$G^\circ(T) = H^\circ - T S^\circ(T) \quad (\text{II.9.14})$$

Afin de pouvoir comparer calculs et données thermochimiques expérimentales, il est nécessaire de considérer l'existence de l'ensemble des conformations du système réel. Dans l'hypothèse d'un équilibre thermique à la température T (distribution de Boltzmann) la fraction molaire  $x_i$  de chaque conformère est donnée par :

$$x_i = \frac{e^{-\frac{G_{T,i}^\circ}{RT}}}{\sum_1^N e^{-\frac{G_{T,i}^\circ}{RT}}} \quad (\text{II.9.15})$$

où N est le nombre de conformères considérés. Ainsi l'enthalpie moyenne d'une mole de ce mélange de conformères peut être déterminée à partir de l'expression :

$$[H^\circ_T]_{moyenne} = \sum_1^N x_i H_{T,i}^\circ \quad (\text{II.9.16})$$

Il en est de même pour l'énergie libre de Gibbs :

$$[G^\circ_T]_{moyenne} = \sum_1^N x_i G_{T,i}^\circ \quad (\text{II.9.17})$$

Pour le calcul de  $[G^\circ_T]_{moyenne}$ , il faut noter que l'entropie moyenne inclut un terme d'entropie de mélange qui correspond au second terme de l'équation :

$$[S^\circ_T]_{moyenne} = \sum_1^N x_i S_{T,i}^\circ - R \sum_1^N x_i \ln x_i \quad (\text{II.9.18})$$

## II.9.2. La méthode NCI

La méthode NCI est une méthode d'analyse, utilisant la densité électronique du système  $\rho$ , et permet d'identifier la nature des liaisons non-covalentes intervenant dans la structure du système étudié.[110-112] Afin d'analyser la nature des interactions, une fonction scalaire a été introduite, il s'agit du gradient réduit de la densité  $s(\rho)$  dépendant à la fois de  $\rho$  et de son gradient :

$$s(\rho) = \frac{|\nabla\rho|}{2(3\pi^2)^{\frac{1}{3}}\rho^{\frac{4}{3}}} \quad (\text{II.9.19})$$

Les interactions non-covalentes correspondent à des zones de l'espace de faibles densités électroniques. Dans ce cas, deux situations sont possibles : la première étant lorsque  $s(\rho) \rightarrow \infty$  pour  $\rho$  prédominant sur  $|\nabla\rho|$  tel que  $\rho \rightarrow 0$ , cela correspond à une région éloignée du système. Le deuxième cas possible est lorsque  $|\nabla\rho| \rightarrow 0$  est prédominant et donc  $s(\rho) \rightarrow 0$ . En général, l'étude du signe du Laplacien de la densité électronique  $|\nabla^2\rho|$  permet l'identification des interactions à électrons partagés ou non, mais ce n'est pas le cas avec la méthode NCI. Ce dernier est décomposé en la somme de trois valeurs propres de la matrice Hessian de la densité électronique selon les trois axes principaux de la variation maximale:

$$\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3 \quad (\text{II.9.20})$$

$\lambda_2$  caractérise la variation de la densité dans un plan perpendiculaire à l'axe internucléaire et elle peut être positive ou négative. C'est cette grandeur qui est utilisée dans la méthode NCI. Si cette variation (i) s'accumule et  $\lambda_2 < 0$ , alors les interactions seront de types liaisons hydrogène ; (ii) diminue et  $\lambda_2 > 0$ , alors il y aura répulsion stérique et (iii) est négligeable ( $\lambda_2 \sim 0$ ), alors les interactions seront de type van der Waals. La fonction  $s(\rho)$  permet d'obtenir des isosurfaces dont les faibles interactions sont visualisables. Pour les visualiser, nous avons utilisé le programme NCIPLOT Version 1.0.[110,111] Une échelle de couleur permet de les différencier : les liaisons hydrogène seront représentées en bleu, les interactions de type van der Waals seront en vert et les interactions type répulsion seront indiquées en rouge.

### II.9.3. Etude de la densité de spin

Pour les systèmes à couches ouvertes, les densités de spin  $\alpha$  et  $\beta$  ne sont pas identiques et sont calculées séparément. La différence entre ces deux densités de spin permet de remonter au(x) spins non apparié(s). Au cours de nos travaux, nous avons obtenu ces densités de spin de deux façons différentes.

#### II.9.3.1. Densité de spin SCF

Les densités de spin des méthodes *ab initio* et DFT ont été extraites de la procédure SCF. Elles sont calculées comme étant la différence entre les densités de spin  $\alpha$  et  $\beta$  :

$$\rho^{spin}(r) = \rho^{\alpha}(r) - \rho^{\beta}(r) = \sum_{i=1}^n |\phi_i^{\alpha}(r)|^2 - \sum_{i=1}^n |\phi_i^{\beta}(r)|^2 \quad (\text{II.9.21})$$

avec  $\phi_i^{\alpha}$  fonctions d'espace correspondant à l'équation. Des fichiers cubes sont générés pour extraire les densités de spin et elles sont visualisées avec le logiciel gaussView 3.090.

#### II.9.3.2. Densité de spin atomique

La densité de spin atomique a été obtenue dans le cadre de l'analyse de population naturelle (NPA).[114] La densité de spin atomique portée par un atome A est calculée par l'équation suivante :

$$(\rho^{spin})_A = NPA(\beta)_A - NPA(\alpha)_A \quad (\text{II.9.22})$$

avec

$$NPA(\beta)_A = \frac{q(A)_{nuc}}{2} - q(A)_{\beta} \quad (\text{II.9.23})$$

et

$$NPA(\alpha)_A = \frac{q(A)_{nuc}}{2} - q(A)_{\alpha} \quad (\text{II.9.24})$$

$q(A)_{nuc}$  correspond à la charge nucléaire,  $q(A)_{\beta}$  et  $q(A)_{\alpha}$  sont les charges naturelles de A en  $\beta$  et  $\alpha$  respectivement.

### ***II.9.4. L'analyse de décomposition en énergie***

L'analyse de décomposition en énergie (EDA)[115-117] est une méthode quantitative permettant de décrire les différents types d'interactions intervenant dans le système. Cette analyse décompose l'énergie d'interaction  $\Delta E_{int}$  entre le système étudié et ces différents fragments à la géométrie du système en trois termes :

$$\Delta E_{int} = \Delta E_{elect} + \Delta E_{Pauli} + \Delta E_{orb} \quad (\text{II.9.25})$$

Le premier terme représente l'interaction électrostatique classique entre les différentes distributions de charges non perturbées des fragments ; ainsi la densité globale est la somme des densités des fragments. La répulsion de Pauli correspond au second terme. Le dernier terme est l'énergie d'interaction orbitale stabilisante. Il est possible d'ajouter un terme énergétique de dispersion provenant de la correction de Grimme définie dans la partie 2.1.7 ; l'équation devient alors :

$$\Delta E_{int} = \Delta E_{elect} + \Delta E_{Pauli} + \Delta E_{orb} + \Delta E_{disp} \quad (\text{II.9.26})$$

Nous avons utilisé cette analyse implémentée dans le programme ADF 2013.01a.

### ***II.9.5. Erreur de superposition de base***

L'énergie d'interaction entre un monomère X et un monomère Y est couramment calculée de la façon suivante :

$$\Delta E_{int}(XY) = E_{XY}^{XY}(XY) - [E_X^X(X) + E_Y^Y(Y)] \quad (\text{II.9.27})$$

L'indice correspond à la géométrie et l'exposant à la base utilisée. L'énergie du dimère est calculée dans la base du dimère et les énergies des monomères sont calculées dans leur propre base et soit dans la géométrie du complexe pour les fragments, soit dans leur propre géométrie. Les calculs de chimie quantique sur des systèmes de taille moyenne se font avec une base finie. Lors du calcul de cette énergie d'interaction, il y a souvent des interactions artificielles stabilisantes qui surestiment cette énergie. Cette erreur artificielle se nomme

erreur de superposition de base (*Basis Set Superposition Error*), notée BSSE. Elle provient de l'utilisation d'une base trop petite et donc le fragment X utilise les fonctions de base de Y pour compenser une base incomplète. Afin de remédier à ce problème, il est possible d'utiliser la correction « counterpoise » de Boys et Bernadi,[118] notée CP et implémentée dans Gaussian 09, en calculant l'énergie du dimère et des deux monomères dans la base du dimère :

$$\Delta E_{BSSE}(X) = E_X^{XY}(X) - E_X^X(X) \quad (\text{II.9.28})$$

$$\Delta E_{BSSE}(Y) = E_Y^{XY}(Y) - E_Y^Y(Y) \quad (\text{II.9.29})$$

Ainsi l'énergie d'interaction avec la correction de BSSE CP est calculée de la façon suivante :

$$\Delta E_{int}^{CP}(XY) = E_{XY}^{XY}(XY) - [E_X^{XY}(X) + E_Y^{XY}(Y)] \quad (\text{II.9.30})$$

La BSSE est faible pour des bases assez étendues en DFT, mais elle est plus importante avec les méthodes post Hartree-Fock. Ainsi, la correction BSSE CP a été utilisée dans le chapitre III partie 1 lors des calculs de données thermochimiques d'hydratation des acides aminés acétylés avec la méthode post Hartree-Fock MP2 associée à la base d'orbitales atomiques 6-311++G(2d,2p), mais pas en DFT.

## II.10. Bibliographie

- [1] Jensen, F. Introduction to Computational Chemistry, 2<sup>nd</sup> Edition, Chichester, *John Wiley & Sons Ltd*, **2007**, 620 pages.
- [2] Atkins, P.; Friedman, R. Molecular Quantum Mechanics, 4<sup>th</sup> Edition, Oxford, *Oxford University Press*, **2005**, 588 pages.
- [3] Koch, W.; Holthausen, M. C. A Chemist's Guide to density Functional Theory, 2<sup>nd</sup> Edition, Weinheim, *Wiley-VCH*, **2002**, 313 pages.
- [4] Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, 87, 5968-5975.
- [5] Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, 98, 7029-7039.
- [6] Krylov, A. I. *Ann. Rev. Phys. Chem.* **2008**, 59, 433-462.
- [7] Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, 22, 1127-1135.
- [8] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, 111, 10439-10452.
- [9] Karton, A.; Gruzman, D.; Martin, J. M. L. *J. Phys. Chem. A* **2009**, 113, 8434-8447.
- [10] Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, 4, 75-85.
- [11] Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. *J. Chem. Phys.* **2008**, 129, 194111-1-14.
- [12] Thakkar, A. J.; McCarthy, S. P. *J. Chem. Phys.* **2009**, 131, 134109-1-12.
- [13] Riley, K. E.; Op't Holt, B. T.; Merz Jr, K. M. *J. Chem. Theory Comput.* **2007**, 3, 407-433.
- [14] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, 109, 5656-5667.
- [15] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, 1, 415-432.
- [16] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, 110, 5121-5129.
- [17] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, 3, 289-300.
- [18] Riffet, V.; Frison, G.; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2011**, 13, 18561-18580.
- [19] Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, 45, 566-569.
- [20] Hedin, L.; Lundqvist, B. I. *J. Phys. C: Solid St. Phys.* **1971**, 4, 2064-2083.
- [21] Vosko, S. J.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 1200-1211.
- [22] Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, 23, 5048-5079.
- [23] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, 45, 13244-13249.
- [24] Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098-3100.
- [25] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865-3868.
- [26] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785-789.
- [27] Perdew, J. P. *Phys. Rev. B* **1986**, 33, 8822-8824.
- [28] Perdew, J. P. in *Electronic Structure of Solids '91*, Berlin, Akademie Verlag, **1991**, 249 pages.
- [29] Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, 109, 400-410.
- [30] Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Phys. Rev. Lett.* **1999**, 82, 2544-2547.
- [31] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, 91, 146401-146401.
- [32] Becke, A. D. *J. Chem. Phys.* **1993**, 98, 1372-1377.
- [33] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215-241.
- [34] Leininger, T.; Stoll, H.; Werner, H.; Savin, A. *Chem. Phys. Lett.* **1997**, 275, 151-160.
- [35] Gerber, I. C.; Ángyán, J. G. *Chem. Phys. Lett.* **2005**, 415, 100-105.
- [36] Gerber, I. C.; Ángyán, J. G. *J. Chem. Phys.* **2007**, 127, 054101-1-9.
- [37] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, 393, 51-57.
- [38] Henderson, T. M.; Izmaylov, A. F.; Scuseria, G. E.; Savin, A. *J. Chem. Phys.* **2007**, 127, 221103 1-4.
- [39] Henderson, T. M.; Izmaylov, A. F.; Scuseria, G. E.; Savin, A. *J. Chem. Theory Comput.* **2008**, 4, 1254-1262.

- [40] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463-1473.
- [41] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104-1-19.
- [42] Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005-1-4.
- [43] Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997-1000.
- [44] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 184108-1-12.
- [45] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105-1-8.
- [46] Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1996**, *104*, 2598-2619.
- [47] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622-5629.
- [48] Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, *93*, 2537-2545.
- [49] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221-7230.
- [50] Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293-1298.
- [51] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063-1079.
- [52] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764-7776.
- [53] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703-4709.
- [54] Baboul, A. G.; Curtiss, L. A.; Redfern, P. C. *J. Chem. Phys.* **1999**, *110*, 7650-7657.
- [55] Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173-1213.
- [56] Voityuk, A. A.; Rösch, N. *J. Phys. Chem. A* **2000**, *104*, 4089-4094.
- [57] Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616-626.
- [58] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.
- [59] Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344-352.
- [60] Korth, M. *J. Chem. Theory Comput.* **2010**, *6*, 3808-3816.
- [61] Stewart, J. J. P. *J. Mol. Model.* **2013**, *19*, 1-32.
- [62] Cornell, W. D.; Cieplak, P.; Bayly, C. B.; Gould, I. R.; Jr Merz, K. M.; Fergusen, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [63] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187-217.
- [64] MacKerell, A. D.; Jr.; Brooks, B.; Brooks, C. L.; III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program; In Schleyer, P.v.R.; et al. *The Encyclopedia of Computational Chemistry 1*: John Wiley & Sons: Chichester, 1998.
- [65] Allinger, N. L.; Tribble, M. T.; Miller, M. A.; Wertz, D. H. *J. Am. Chem. Soc.* **1977**, *99*, 8127-8134.
- [66] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8566.
- [67] Allinger, N. L.; Chen, K. S.; Lii, J. H. *J. Comput. Chem.* **1996**, *17*, 642-668.
- [68] Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.
- [69] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933-5947.
- [70] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
- [71] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, *9*, 4046-4063.



- [73] Kaminsky, J.; Jensen, F.; *J. Chem. Theory Comput.* **2007**, *3*, 1774-1788.
- [74] Riffet, V.; Frison, G.; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18561-18580.
- [75] Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *J. Phys. Chem. B* **2012**, *116*, 483-490.
- [76] Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450-3462.
- [77] Semrouni, D.; Balaj, O. P.; Calvo, F.; Correia, C.; Clavaguéra, C.; Ohanessian, G. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 728-738.
- [78] Martens, J. K.; Compagnon, I.; Nicol, E.; McMahon, T. B.; Clavaguéra, C.; Ohanessian, G. *J. Phys. Chem. Lett.* **2012**, *3*, 3320-3324.
- [79] Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233-239.
- [80] Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047-1064.
- [81] Tholé, B. T. *Chem. Phys.* **1981**, *59*, 341-350.
- [82] Wu, J. C.; Chattree, G.; Ren, P. *Theor. Chem. Acc.* **2012**, *131*, 1138-1148.
- [83] Xiang, J. X.; Ponder, J. W. *J. Comput. Chem.* **2013**, *34*, 739-749.
- [84] Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111-5116.
- [85] Michaux, C.; Wouters, J.; Jacquemin, D.; Perpète, E. A. *Chem. Phys. Lett.* **2007**, *445*, 57-61.
- [86] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Chem. Phys. B* **2008**, *112*, 2430-2438.
- [87] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Chem. Phys. B* **2008**, *112*, 9896-9902.
- [88] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Chem. Phys. B* **2008**, *112*, 7702-7705.
- [89] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 632-638.
- [90] Campo-Cacharrón, A.; Cabaleiro-Lago, E. M.; Rodríguez-Otero J. *Theor. Chem. Acc.* **2012**, *131*, 1290-1302.
- [91] Jacquemin, D.; Michaux, C.; Perpète, E. A.; Frison, G. *J. Chem. Phys. B* **2011**, *115*, 3604-3613.
- [92] Yao, Y.; Chen, D.; Zhang, S.; Li, Y.; Tu, P.; Liu, B.; Dong, M. *J. Chem. Phys. B* **2011**, *115*, 6213-6221.
- [93] Aikens, C. M.; Gordon, M. S. *J. Am. Chem. Soc.* **2006**, *128*, 12835-12850.
- [94] Calvo, F.; Douady, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3404-3414.
- [95] Lemoff, A. S.; Bush, M. F.; Wu, C.; Williams, E. R. *J. Am. Chem. Soc.* **2005**, *127*, 10276-10286.
- [96] Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Chem. Phys. A* **2005**, *109*, 1903-1910.
- [97] Kamariotis, A.; Boyarkin, O.; Mercier, S.; Beck, R.; Bush, M.; Williams, E.; Rizzo, T. J. *Am. Chem. Soc.* **2006**, *128*, 905-916.
- [98] Hratchian, H. P.; Schlegel, H. B. *J. Chem. Phys.* **2004**, *120*, 9918-9924.
- [99] Hratchian, H. P.; Schlegel, H. B. *Theory and Applications of Computational Chemistry: The First 40 Years*, Amsterdam, Elsevier, **2005**, 195-249.
- [100] Hratchian, H. P.; Schlegel, H. B. *J. Chem. Theory Comput.* **2005**, *1*, 61-69.
- [101] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.
- [102] van der Kamp, M. W.; Mulholland, A. J. *Biochem.* **2013**, *52*, 2708-2728.
- [103] Monari, A.; Rivail, J. L.; Assfeld, X. *Acc. Chem. Res.* **2013**, *46*, 596-603.
- [104] Salahub, D. R.; de la Lande, A.; Goursot, A.; Zhang, R.; Zhang, Y. *Struct. Bond.* **2013**, *150*, 1-64.
- [105] Groenhof, G. *Methods in molecular biology* **2013**, *924*, 43-66.
- [106] Steinbrecher, T.; Marcus, E. *Methods in molecular biology* **2013**, *924*, 91-124.

- [107] Thiel ; W.; Hummer, G. *Nature* **2013**, 504, 97-98.
- [108] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, 102, 1995-2001.
- [109] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, 24, 669-681.
- [110] Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, 132, 6498-6506.
- [111] Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J. P.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2011**, 7, 625-632.
- [112] Contreras-García, J.; Yang, W. *J. Phys. Chem. A* **2011**, 115, 12983-12990.
- [113] Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, 14.1, 33-38.
- [114] Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, 83, 735-746.
- [115] Morokuma, K. *J. Chem. Phys.* **1971**, 55, 1236-1244.
- [116] Ziegler, T.; Rauk, A. *Theor. Chem. Acta* **1977**, 46, 1-10.
- [117] von Hopffgarten, M.; Frenking, G. *Comput. Mol. Sci.* **2012**, 2, 43-62.
- [118] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, 19, 553-566.



# Chapitre III : Effets des interactions non-covalentes sur les propriétés structurales et énergétiques



Partie I : Influences réciproques des  
interactions non-covalentes des acides  
aminés acétylés micro-hydratés



### III.1.1. Introduction

Dans le chapitre I, nous avons vu l'intérêt d'étudier la solvation progressive des acides aminés et des petits peptides. Par ajout contrôlé de molécules de  $\text{H}_2\text{O}$ , elle permet notamment d'obtenir diverses informations structurales et thermochimiques du système considéré. Les processus de solvation sont plus clairs lorsque les données expérimentales sont associées à des structures moléculaires. Dans ce travail, une étude théorique sur les acides aminés (X)  $\text{N}\alpha$ -acétylés (NA) protonés solvatés par  $n$  molécules de  $\text{H}_2\text{O}$ , noté  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  avec  $n = 0$  à 3, a été investi pour les acides aminés basiques afin de nous permettre d'étudier l'effet de la nature du site chargé sur l'hydratation. Les modifications des propriétés thermochimiques et structurales de ce type de molécules dû à l'hydratation progressive ont également été étudiées.

Un acide aminé  $\text{N}\alpha$ -acétylé est un acide aminé dans lequel un groupement acyle est fixé sur l'atome d'azote directement lié au carbone  $\alpha$  ( $\text{C}\alpha$ ). Ce type de composé possède une liaison peptidique. La présence de cette liaison fait de ces systèmes de petits modèles d'études pour les peptides. Leur forme générale est représentée dans la Figure III.1.1.

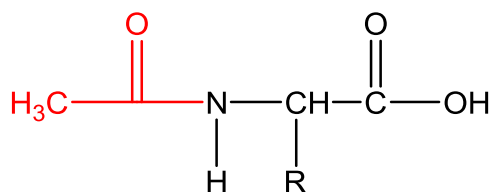


Figure III.1.1. Représentation d'un acide aminé  $\text{N}\alpha$ -acétylé.

Le site de protonation favorable des acides aminés  $\text{N}\alpha$ -acétylés est généralement le groupement acyle  $\text{H}_3\text{CCO}$ , [1,2] excepté pour les acides aminés qui possèdent des chaînes latérales basiques, c'est-à-dire les groupements guanidine ( $\text{R}-\text{CH}_4\text{N}_3$ ), imidazole ( $\text{R}-\text{C}_3\text{H}_3\text{N}_2$ ) et amine ( $\text{R}-\text{NH}_2$ ) présents respectivement dans l'arginine (Arg, R), l'histidine (His, H) et la lysine (Lys, K). [1,2] À pH physiologique, les chaînes latérales de la lysine et de l'arginine sont sous forme protonées. Cela n'est pas systématiquement le cas du groupement imidazole de l'histidine, qui n'est protoné que pour les pH physiologiques inférieurs ou égaux à 6. Toutefois, ce dernier peut être un acteur majeur dans les processus de transfert de proton. Dans le cadre de nos travaux, nous avons étudié les  $\text{N}\alpha$ -acetyl- histidine, lysine et arginine, notés  $[\text{NAH} + \text{H}]^+$ ,  $[\text{NAK} + \text{H}]^+$  et  $[\text{NAR} + \text{H}]^+$ , respectivement (Figure III.1.2).



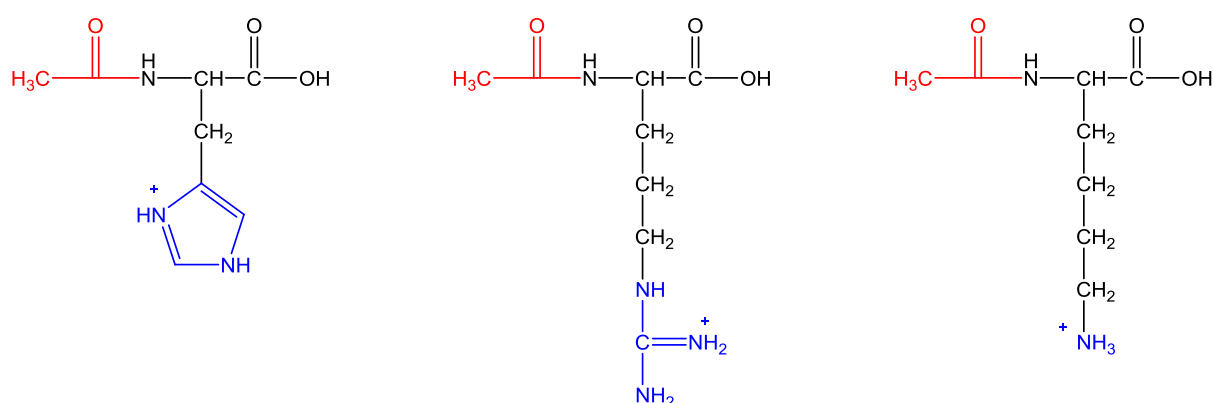
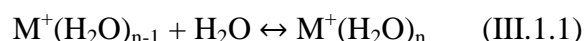


Figure III.1.2. Représentation de  $[\text{NAH} + \text{H}]^+$ ,  $[\text{NAR} + \text{H}]^+$  et  $[\text{NAK} + \text{H}]^+$ . En rouge, les groupements acyles et en bleu les sites chargés.

L'hydratation d'une molécule est influencée par son état de charge, comme le montrent des travaux antérieurs.[3-5] Les premières molécules d'eau interagissent plus fortement avec les molécules de plus haut états de charges. Par exemple, l'enthalpie de mono-hydratation ( $\Delta H_1$ ) de la neurotensine d'état de charge  $+z$ , un neuropeptide composé de 13 acides aminés (ELYENKPRRPYIL), est égale à -38.5, -43.1 et -61.9 kJ mol<sup>-1</sup> pour un état de charge de +1, +2 et +3, respectivement.[4] Dans cette partie, nous nous sommes systématiquement limité à l'état de charge +1.

La mesure des données thermochimiques d'hydratation, notées  $\Delta H_n$ ,  $\Delta S_n$  et  $\Delta G_n$  pour indiquées qu'elles concernent la n<sup>ième</sup> molécule d'eau, est réalisée via la réaction d'équilibre thermique suivante :



qui a lieu en phase gazeuse dans des sources d'ion à haute pression ou dans des cellules de dérive à eau sous pression dans un spectromètre de masse. Des mesures de ces données ont été réalisées pour des molécules possédant les groupements ammonium ( $\text{NH}_3^+$ ), guanidinium ( $\text{CH}_6\text{N}_3^+$ ) et imidazolium ( $\text{C}_3\text{H}_5\text{N}_2^+$ ). Les valeurs obtenues pour les systèmes d'intérêts pour notre étude sont regroupés dans le Tableau III.1.1. Il s'agit de molécules possédant l'un des trois sites chargés (ammonium, guanidinium et imidazolium), un ou plusieurs carbonyles et (ou) qui sont N-acétylé. Certains de ces systèmes sont des modèles des  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  (par exemple  $\text{CH}_3\text{NH}_3^+(\text{H}_2\text{O})$  et  $\text{C}_3\text{H}_5\text{N}_2^+(\text{H}_2\text{O})$  correspondent respectivement à une portion de la chaîne latérale de  $[\text{NAK} + \text{H}]^+$  et  $[\text{NAH} + \text{H}]^+$  monohydratés). Inversement, nos

systèmes d'études sont des modèles de peptides plus étendus tels que  $[\text{Ac-AAAAK} + \text{H}^+]$ ,  $[\text{Ac-AARAA} + \text{H}^+]$  ou  $[\text{Ac-A}_{15}\text{H} + \text{H}^+]$ . Des comparaisons entre ces systèmes permettront de rechercher les liens entre structure géométrique, présence d'interactions non-covalentes, site d'hydratation et données thermochimiques.

Pour l'ensemble des données regroupées dans le Tableau III.1.1, il est probant que l'augmentation du nombre de molécules d'eau interagissant avec ces systèmes induit une diminution de l'énergie d'interaction en termes de  $\Delta H_n$  et  $\Delta G_n$ , effet bien connu dans la littérature. Toutefois, les tendances de l'entropie d'hydratation ( $\Delta S_n$ ) sont moins nettes et ne semblent pas pouvoir être généralisées. Par exemple, dans le cas du groupe ammonium, on note que  $\Delta S_n(\text{CH}_3\text{NH}_3^+)$  augmente avec  $n$  tandis que  $\Delta S_n(\text{LysH}^+)$  diminue et qu'il n'y a pas d'évolution distinct pour  $\Delta S_n(\text{N}\alpha\text{-AcLysH}^+)$ .

Dans la littérature, les valeurs expérimentales pour une même molécule peuvent être sensiblement différentes. Par exemple, l'enthalpie de mono-hydratation du méthylammonium a été mesuré par différentes équipes, et les valeurs obtenues vont de  $-62.3 \pm 4.6$ , [6] à  $-70.3$ [7] et  $-78.7$ [8]  $\text{kJ mol}^{-1}$ . De ce fait, il peut s'avérer difficile par exemple de conclure sur la force de l'interaction (équivalente ou plus forte) entre  $\text{CH}_3\text{NH}_3^+ \cdots \text{OH}_2$  et  $\text{C}_3\text{H}_5\text{N}_2^+ \cdots \text{OH}_2$  ( $-61.9 \text{ kJ mol}^{-1}$ ). Ainsi, dans le but de comprendre les tendances observées des valeurs thermochimiques d'hydratation des  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$ , nous avons choisi de nous baser dans un premier temps sur les valeurs thermochimiques d'hydratation calculées de tous les systèmes modèles utilisés, avant de les confronter aux données expérimentales dans le cas de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$ , seul système étudié pour lequel elles sont disponibles.

Avant d'étudier les conséquences de la micro-hydratation sur les propriétés structurales et thermochimiques des  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  (partie III.1.3), l'identification des structures moléculaires est primordiale. Pour cela nous avons établi, en deux étapes, une stratégie d'exploration de la surface d'énergie potentielle qui est exposée en détails dans les parties III.1.2.1 et III.1.2.2. La confrontation des données thermochimiques théoriques et expérimentales de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  validera notre approche d'exploration de la surface d'énergie potentielle.

Tableau III.1.1. Données thermochimiques d'hydratation à 298 K ( $\Delta H^\circ_n$ ,  $\Delta S^\circ_n$  et  $\Delta G^\circ_n$ ) concernant les groupements ammonium, guanidinium et imidazolium, reportées dans la littérature.

	n	$\Delta H^\circ_n$ <sup>a</sup>	$\Delta S^\circ_n$ <sup>b</sup>	$\Delta G^\circ_n$ <sup>a</sup>	Référence		n	$\Delta H^\circ_n$ <sup>a</sup>	$\Delta S^\circ_n$ <sup>b</sup>	$\Delta G^\circ_n$ <sup>a</sup>	Référence
<b>ammonium</b>											
CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	1	-62.3±4.6	-83.3±13.0	-38.0±8.4	[6]	N-AcLysH <sup>+</sup>	1	/	/	-24.3	[10]
		-70.3	/	/	[7]						
		-78.7	/	/	[8]	Nα-AcLysH <sup>tc</sup>	1	-44.3±1.3	-73.6±4.2	-22.4±2.6	[12]
	2	-54.8±2.5	-90.0±7.9	-28.0±5.0	[6]		2	-35.1±1.3	-61.1±4.2	-16.9±2.6	[12]
		-61.1	/	/	[7]		3	-34.7±1.3	-69.9±4.2	-13.9±2.6	[12]
		-61.1	/	/	[8]		4	-30.1±1.3	-61.5±4.2	-11.8±2.6	[12]
	3	-51.0±2.5	-102.5±7.9	-20.5±5.0	[6]	Nε-AcLysH <sup>+</sup>	1	-42.3±1.3	-69.9±4.2	-21.5±2.6	[12]
		-51.5	/	/	[7]		2	-37.2±1.3	-66.9±4.2	-17.3±2.6	[12]
		-51.9	/	/	[8]		3	-32.2±1.3	-66.5±4.2	-12.4±2.6	[12]
	4	-43.1	/	/	[7]						
	5	-38.0	/	/	[7]	Ac-AAKAA+H <sup>+</sup>	1	-35.6±2.9	-74.9±8.4	-13.3±5.4	[12]
	6	-35.6	/	/	[7]		2	-31.0±2.9	-58.2±8.4	-13.7±5.4	[12]
n-C <sub>5</sub> H <sub>11</sub> NH <sub>3</sub> <sup>+</sup>	1	-61.9±0.8	-97.1±2.9	-33.0±1.7	[6]	Ac-AAAAK+H <sup>+</sup>	1	-28.9±2.9	-56.5±8.4	-12.1±5.4	[12]
	2	-56.9±1.7	-109±5.9	-24.3±3.8	[6]						
	3	-45.2±4.2	-93.3±15.1	-17.6±9.2	[6]	Ac-KA <sub>15</sub> +H <sup>+</sup>	1	-48.9±2.3	-131±10	-9.9±5.3	[3]
GlyH <sup>+</sup>	1	-64.4±2.9	-85.3±6.7	-39.3±5.0	[9]	Ac-KA <sub>20</sub> +H <sup>+</sup>	1	-47.0±3.5	-114±15	-13.0±8.0	[3]
		/	/	-40.6	[10]						
	2	-58.2±1.7	-93.7±5.9	-30.5±4.2	[9]						
		/	/	-30.1	[10]						
	3	-53.6±1.3	-103.8±3.8	-23.0±2.5	[9]						
		/	/	-22.6	[10]						
LysH <sup>+</sup>	4	-44.3±2.5	-92.9±8.4	-17.1±5.0	[9]						
	1	-45.6	-84.5	-20.4	[11]						
		/	/	-21.8	[10]						
	2	-40.2	-80.7	-16.1	[11]						
	3	-33.0	-70.3	-12.0	[11]						
	4	-28.0	-59.0	-10.4	[11]						
	5	-24.3	-46.9	-10.3	[11]						

<sup>a</sup> kJ mol<sup>-1</sup>. <sup>b</sup> J.K<sup>-1</sup>.mol<sup>-1</sup>. <sup>c</sup> correspond à [NAK + H]<sup>+</sup>.

Tableau III.1.1 – suite. Données thermochimiques d'hydratation à 298 K ( $\Delta H_n^\circ$ ,  $\Delta S_n^\circ$  et  $\Delta G_n^\circ$ ) concernant les groupements ammonium, guanidinium et imidazolium, reportées dans la littérature.

	n	$\Delta H_n^\circ$ <sup>a</sup>	$\Delta S_n^\circ$ <sup>b</sup>	$\Delta G_n^\circ$ <sup>a</sup>	Référence		n	$\Delta H_n^\circ$ <sup>a</sup>	$\Delta S_n^\circ$ <sup>b</sup>	$\Delta G_n^\circ$ <sup>a</sup>	Référence
<b>guanidinium</b>						<b>imidazolium</b>					
ArgH <sup>+</sup>	1	-43.1	-82.4	-18.5	[11]	Imidazolium	1	-61.9	-99.6	-32.2	[14]
		-38.0	-64.4	-18.8	[13]						
	2	-36.4	-75.7	-13.8	[11]	Ac-A <sub>15</sub> H+H <sup>+</sup>	1	-62.7±3.0	-187±13.0	-7.0±6.9 (298K)	[3]
	3	-31.4	-66.5	-11.6	[11]						
	4	-27.2	-55.6	-10.6	[11]						
	5	-27.2	-62.3	-8.6	[11]						
RAAAA+H <sup>+</sup>	1	-38.9±1.3	-84.5±4.2	-13.7±2.6	[4]						
	2	-32.6±1.3	-73.6±4.2	-10.7±2.6	[4]						
	3	-29.7±1.3	-61.9±4.2	-11.2±2.6	[4]						
AARAA+H <sup>+</sup>	1	-42.7±1.3	-96.2±4.2	-14.0±2.6	[4]						
	2	-35.1±1.3	-75.3±4.2	-12.7±2.6	[4]						
Ac-AARAA+H <sup>+</sup>	1	-39.7±1.3	-87.9±4.2	-13.5±2.6	[4]						
	2	-33.9±1.3	-75.3±4.2	-11.5±2.6	[4]						

<sup>a</sup> kJ mol<sup>-1</sup>. <sup>b</sup> J.K<sup>-1</sup>.mol<sup>-1</sup>.

## III.1.2. Exploration de la surface d'énergie potentielle

### *III.1.2.1. Couplage des approches hiérarchique et arbre de Darwin*

L'exploration de la surface d'énergie potentielle des  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  est une étape clé avant d'étudier les effets de la micro-hydratation sur les propriétés structurales et thermochimiques. Notre premier objectif était d'obtenir une population de conformères de basse énergie la plus exhaustive possible. Une étude antérieure réalisée sur la micro-hydratation de la glycine protonée a montré que l'approche hiérarchique et celle basée sur la construction d'un arbre de Darwin (approches expliquées au chapitre II partie 6.1.2) ne permettent pas à elles seules d'identifier l'ensemble des minima sur la surface d'énergie potentielle.[15] Cependant, la réalisation de ces deux approches en parallèle permet d'obtenir une population de conformères plus complète. En effet, comme le montre l'exemple de  $[\text{Gly} + \text{H}]^+(\text{H}_2\text{O})_2$ , 23 structures peuvent être obtenues dans la fenêtre d'enthalpie libre relative 0-10 kcal/mol au niveau CP-MP2/6-311++G(d,p), mais seules 19 étaient obtenues par l'approche hiérarchique. De même, 5 structures manquaient (dont 3 n'avaient pas été recherchées) en se limitant à la construction d'un arbre type Darwin. Pourtant, la glycine protonée semble un cas relativement « simple », car peu flexible (elle ne possède que trois conformations dans la génération  $\text{G}_0$ ) et ne présentant que deux sites donneurs de liaison hydrogène (OH et  $\text{NH}_3$ ) et un site accepteur de liaison hydrogène (CO). Sur la base de ces constatations, à partir de ces approches hiérarchique et de construction d'un arbre de Darwin, nous avons mis au point et testé une stratégie couplant efficacement ces deux approches pour générer la population de conformères des systèmes  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  la plus complète possible. L'idée générale était non pas de faire ces deux approches en parallèle, ce qui multiplie par deux le travail à effectuer, mais de les réaliser de manière séquentielle : une exploration complète de la surface d'énergie potentielle avec AMOEBA, suivie d'une ré-optimisation de la géométrie au niveau DFT des minima situés dans une gamme donnée d'énergie, permet d'obtenir un premier lot de minima pour chaque état d'hydratation. A partir de ces minima, un arbre de Darwin est construit et les structures manquantes (dans les générations précédentes et/ou suivantes) sont recherchées manuellement. Dans un premier temps, nous avons testé cette stratégie pour le

système  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$ , qui est le système a priori le moins flexible des trois que nous souhaitons étudier car sa chaîne latérale ne comporte qu'un groupe  $\text{CH}_2$ . Ces résultats ont fait l'objet de la rédaction d'un article, qui doit être soumis prochainement pour publication, article qui est reproduit intégralement ci-après.

Tout d'abord, voici quelques précisions sur une partie de la procédure pour l'exploration de la surface d'énergie potentielle : nous avons construit différentes structures comme point de départ des MC, saut de bassins et dynamiques moléculaires, présentant différents types d'interactions et (ou) sites de fixation pour les molécules  $\text{H}_2\text{O}$ . Pour une même structure de départ, nous avons lancé en parallèle plusieurs dynamiques à des températures différentes (300, 500, 750 et 1000 K) avec des temps de dynamique de 0.16 et 4.0 ns pour lesquels les structures ont été sauvegardées tous les 0.16 et 4.0 ps, respectivement. Toutes les structures générées lors de ces dynamiques moléculaires ont été considérées et optimisées avec AMOEBA.

Dans le cadre du MC et du saut de bassins, la recherche de minima est effectuée par des mouvements de torsions. Nous avons gelé la liaison peptidique en conformation *trans* car il est connu que cette conformation est favorisée du fait des interactions stériques entre les chaînes latérales liées au carbone  $\alpha$ , sauf dans le cas de la proline.[16] Lors du MC, nous avons considéré un nombre totale de pas de 1000 ps une température de 500 K et lors du saut de bassins, la fenêtre d'énergie considérée était de  $418 \text{ kJ mol}^{-1}$ .

# Microhydration of protonated N $\alpha$ -acetylhistidine: a theoretical approach

Vanessa Riffet, Guy Bouchoux and Gilles Frison\*

Laboratoire de Chimie Moléculaire, Département de Chimie, Ecole Polytechnique and CNRS, 91128 Palaiseau cedex, France.

## Corresponding author:

gilles.frison@polytechnique.org

Tel: +33 (0)1 69 33 48 34

Fax: +33 (0)1 69 33 48 03

## KEYWORDS:

Density functional theory; AMOEBA force field; Non-covalent interaction; Thermochemistry; Microsolvation; Hydration energies.

## ABSTRACT:

Theoretical studies to determine structures, hydration sites and thermochemical quantities are challenging, especially for flexible molecules including several functional groups, which allow different hydration sites. A detailed theoretical study of protonated N $\alpha$ -acetylhistidine ([NAH+H]<sup>+</sup>) hydrated by 0 to 3 water molecules was carried out. Extensive exploration of the corresponding potential energy surfaces was performed through combination of a hierarchical and Darwin tree approaches using polarizable AMOEBA force field and M06 DFT functional. We show that this mixed approach allows exploring, as exhaustively and as accurately as possible, the potential energy surface of the considered systems without loss of conformers during the exploration due to low quality of the intermediate levels of calculation or conformational changes induced by microhydration. Hydration enthalpies and entropies have been computed using both a single conformer and equilibrium populations of conformers at 298 K. In this latter situation, averaged  $\Delta H^{\circ}_{n,average}$  of -58, -55 and -47 kJ mol<sup>-1</sup> have been calculated for n = 1, 2 and 3, respectively. Thermochemical parameters of [NAH+H]<sup>+</sup> and those of model compounds have been compared to understand reciprocal influences of [NAH+H]<sup>+</sup> and water molecules.

## INTRODUCTION

Understanding the tridimensional structure of proteins in bulk water is far from being achieved. No doubt however that protein folding is intimately influenced by intra- and intermolecular interactions, among which hydrophobic forces and hydrogen bonds.<sup>1,2</sup> In this context, studies of solvation by a discrete number of water molecules, i.e. microhydration, allows bridging the gap between isolated and fully solvated molecules. This strategy has been used in a number of experimental and theoretical studies for solutes ranging from inorganic ions to protonated biomolecules such as amino acids or peptides. For instance, complexation thermochemistry for hydration by a discrete number of water molecules of protonated amino acids and peptides has been determined by several mass spectrometry techniques. Indeed, thermal equilibrium of reactions:  $A^+(H_2O)_{n-1} + H_2O \rightarrow A^+(H_2O)_n$ , have been performed into high pressure ion sources<sup>3-8</sup> or into water pressurized drift cells<sup>9-13</sup> thus allowing the corresponding Gibbs free energy change to be determined. In addition, measuring the equilibrium constants as a function of temperature allows the determination of the relevant enthalpy and entropy changes via a van't Hoff type analysis. Blackbody infrared radiative dissociation (BIRD) may provide thermochemical parameters via numerical modeling of dissociation rates of  $A^+(H_2O)_n$  complexes.<sup>14</sup> Similarly, complexation energies may be deduced from modeling of dissociation efficiencies originating from guided ion beam experiments.<sup>15</sup> Beside these experimental source of thermochemical data, infrared multiphoton dissociation (IRMPD) coupled with mass spectrometry brings direct information on the geometry of the complexes  $A^+(H_2O)_n$  due to typical vibrational signatures of C=O, OH and NH stretching modes.<sup>16-18</sup>

Insight into the solvation process is however only possible if experimental informations are coupled with computation of the corresponding molecular structures. From this point of view, difficulties arise when several sites for hydration are in competition thus leading to an increasing number of isomers and conformers in the hydrated ion population. Therefore, the computational approach used to explore the lowest energy part of the potential energy surface of the system should be both accurate and exhaustive. In the recent years, two theoretical approaches have been proven successful to explore the potential energy surface of microhydrated biological molecules and ions. The first strategy, based on the creation of a Darwinian family tree, considers that each generation  $G_n$  of a molecule microsolvated by  $n$  water molecules is determined from previous generation  $G_{n-1}$ . This strategy has been developed<sup>19</sup> and successfully used for small protonated amino acids such as Gly,<sup>20</sup> Ala<sup>21</sup> and Pro<sup>22</sup>. Larger systems like deprotonated Asp and Gln behave in the same way,<sup>23</sup> as well as others molecular systems like cation-phenol.<sup>24</sup> A recent experimental study also confirms that the linear conformation of diprotonated 1,7-diaminoheptane is kept under hydration up to 10 water molecules.<sup>25</sup> However, adding more than 12 water molecules induces a more folded conformation, which means that the most stables conformers of  $G_{12}$  should not be related to the ones of  $G_{10}$ .<sup>25</sup> Other examples where hydration induces change in the protonation site,<sup>26,27</sup> inducing possible isomerization between non-zwitterionic and zwitterionic structures,<sup>28-31</sup> or in the conformational structures,<sup>32</sup> show that this computational strategy remains to be evaluated for flexible systems. Furthermore, this strategy relies on the knowledge of the first generation  $G_0$  which could also be a challenging task for flexible molecules. The second strategy is a hierarchical approach, where computational steps of increasing accuracy are used successively. In most cases, as a first step, a large number of starting geometries are generated by molecular dynamics (MD) and/or Monte Carlo (MC) calculations using force fields. Candidate low-energy structures are then re-optimized using quantum theoretical tools, either directly with state-of-the-art level<sup>33,34</sup> or after one (or several) intermediate level(s) after which the less reliable structures are eliminated<sup>35-38</sup>. One crucial point in this strategy is the ability of force fields, and more generally the used intermediate levels, to explore satisfyingly the potential energy surface both in terms of structural geometries and relative energies. Jensen et al. have shown for

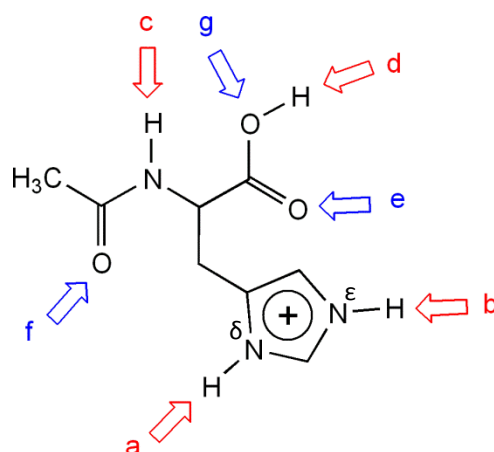


neutral acetyl and N-methyl amino acids<sup>39</sup> the limitation of conventional force fields, whereas the AMOEBA polarizable force field<sup>40,41</sup> gives significantly more reliable structures and energies. Its relevance, due to an efficient treatment of electrostatics and polarization effects, has been confirmed for charged systems such as protonated and deprotonated methionine,<sup>42</sup> Gramicidin S peptide<sup>43</sup> or sodiated oligoglycines.<sup>44</sup>

In a recent paper, it has been shown that despite the relevance of these two approaches, they do not allow to find all the lowest energy conformers.<sup>45</sup> However, these two strategies turn to be complementary as the evolutionary tree approach fills in the missing structures of the hierarchy of scheme one, and reciprocally.

In this work we defined a new way to match these two approaches in order to take advantage of their complementarity without significantly increasing the computational time. The two strategies have not been followed in parallel, as done previously,<sup>45</sup> but sequentially. First, the hierarchical approach is used for each generation to identify the most stable conformers. Then the Darwin tree is built based on all the minima identified. This allows checking the absence of structures for a given family, that is a *children* at any generation relative to one specific conformation observed at the first generation, which can therefore be added manually. This verification should prevent gaps in the hierarchical strategy due to the somehow imprecision of the computational methods used. Inversely, the observation of a new conformer at the  $n^{\text{th}}$  generation, which is not related to any *parents*, allows including a new family to the tree which would have been missed following only the Darwin tree strategy.

N-acetylated aminoacids are simple models systems allowing investigating the microhydration process of molecules containing a peptide bond. In the gas phase, most of the isolated N-acetylated aminoacids protonate on the carbonyl of the acyl moiety.<sup>46,47</sup> Exceptions of this rule are aminoacids containing a basic side chain such as Arg, Lys and His.<sup>46,48</sup> The present study is focused on microhydration of protonated N $\alpha$ -acetylhistidine, noted  $[\text{NAH}+\text{H}]^+$ . Histidine itself has been studied in great detail recently demonstrating the favoured protonation on the imidazole group and the participation of several types of internal hydrogen bonds in the stabilization of neutral, protonated and deprotonated histidine.<sup>49</sup> In protonated N $\alpha$ -acetylhistidine, new possibilities of internal hydrogen bond are obviously opened. Furthermore, this small model peptide presents several sites able to interact with water molecule, among them the COOH-terminal group and the amide and imidazolium NH groups (Figure 1). The result of the competition between fixation of one water molecule at one of these chemical groups or possible insertion into an intramolecular H-bond is a question explored in this study. Last, we propose a set of evaluated theoretical gas-phase hydration thermochemical parameters including the various conformers identified in the 298 K populations of  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$  ( $n=0-3$ ) species.



**Figure 1.** Structure of  $[\text{NAH}+\text{H}]^+$ . Hydrogen bond donor and acceptor sites are indicated with red and blue arrows, respectively.

## METHODOLOGY

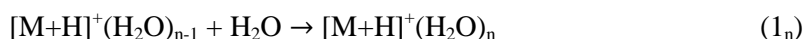
**Computational Details.** As indicated in the introduction, the strategy adopted to explore exhaustively the conformational landscapes of  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 0$  to 3) complexes combines the hierarchical and Darwin tree approaches. The M06 DFT functional is known to perform satisfactorily for thermochemistry and non-covalent interactions calculations,<sup>50</sup> it has therefore been used throughout this study using the double- $\zeta$  quality basis set 6-31+G(d,p), as implemented in the Gaussian 09 package,<sup>51</sup>

(i) For the hierarchical approach, we have used MD and MC simulations, together with basin-hopping conformational scanning algorithm as implemented in the Tinker molecular modeling package<sup>52</sup> associated with AMOEBA polarizable force field. The parameters for the neutral COOH terminations were adjusted using the parameters of the carboxylic group of the aspartic acid which are already available in the AMOEBA parameter set.<sup>45</sup> This exploration has been fully achieved for  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 0$  to 3) to generate a first set of structures. This step was stopped when no new minimum appears at the AMOEBA level following MD, MC or basin-hopping simulation starting from a random conformation. This procedure led to 143, 470, 1076 and 3358 minima for  $n = 0, 1, 2$  and 3, respectively. Then, for  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  and  $\mathbf{G}_3$  generation (i.e. populations of complexes  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1$  to 3)), all structures situated in the 0-20 kJ mol<sup>-1</sup> AMOEBA energy range ( $\Delta E$ ), relative to the most stable conformers were optimized at the M06/6-31+G(d,p) level. For generation  $\mathbf{G}_0$ , which is crucial for the building of a Darwinian tree, a  $\Delta E$  of 0-60 kJ mol<sup>-1</sup> has been considered and the geometry optimization was conducted directly at the higher M06/6-311++G(d,p) level. This generated 59, 37, 44 and 113 conformers for  $n = 0, 1, 2$  and 3, respectively. Vibrational frequencies were calculated at the same DFT level to obtain enthalpies and Gibbs free energies at 298 K. All minima situated in the 0-15 kJ mol<sup>-1</sup> M06/6-31+G(d,p) Gibbs free energy range for  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  and  $\mathbf{G}_3$  are then re-optimized at the M06/6-311++G(d,p) level based on M06/6-31+G(d,p) geometries.

(ii) All the structures obtained at the M06/6-311++G(d,p) level for each generation  $\mathbf{G}_0$ ,  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  and  $\mathbf{G}_3$  were then used to build a Darwinian tree. Examination of the resulting tree allowed us to identify missing structures in a family that are structures not observed at  $\mathbf{G}_n$  population even if related structures were present in other generation  $\mathbf{G}_m$  ( $m \neq n$ ). Furthermore, the Darwinian tree allows us to identify hydrogen bonds donor and acceptor groups available for water molecule fixation on the conformations obtained in the different generation, and to build corresponding missing structures, respectively. The missing structures identified in the preceding step are built manually and optimized at the M06/6-311++G(d,p) level. In considering all generation  $\mathbf{G}_n$ , a total of 209 new geometries have been optimized at this step, among them 29 belong to the 0-15 kJ mol<sup>-1</sup> range. All geometries, relative energies and thermochemical data discussed in this work are then obtained from optimization and vibrational frequencies calculations at the M06/6-311++G(d,p).

Non-covalent interactions are visualized using the NCIPLOT software Version 1.0.<sup>53-54</sup> Figures are shown for SCF densities. NCI surfaces correspond to  $s = 0.5$  au and the color scale is  $-0.05 < \text{sign}(\lambda_2)\rho < +0.05$  au. Blue indicates strong attractive interactions, green indicates van der Waals interactions and red indicates strong repulsive interactions.

**Thermochemistry.** In this study, we have calculated enthalpies ( $\Delta H^\circ_n$ ), entropies ( $\Delta S^\circ_n$ ) and Gibbs free energies ( $\Delta G^\circ_n$ ) of hydration of species  $[\text{M}+\text{H}]^+(\text{H}_2\text{O})_{n-1}$  by considering the reaction (1<sub>n</sub>) in gas phase.



These quantities are estimated in considering one mol of gaseous  $[\text{M}+\text{H}]^+(\text{H}_2\text{O})_n$ ,  $[\text{M}+\text{H}]^+(\text{H}_2\text{O})_{n-1}$  and  $\text{H}_2\text{O}$  in thermal equilibrium at temperature  $T=298\text{K}$ .

In a first approach, each mol of reactant and product is considered to contain only one conformation of generation  $\mathbf{G}_n$  and the related conformation in the parent generation  $\mathbf{G}_{n-1}$  after steps (i) and (ii) as

described in the methodology part. The notation “same conformation” (sc) is used and  $\Delta H^\circ_{n,sc}$ ,  $\Delta S^\circ_{n,sc}$  and  $\Delta G^\circ_{n,sc}$  may be calculated using, respectively equations 2<sub>n</sub>-4<sub>n</sub>:

$$\Delta H^\circ_{n,sc} = H^\circ_{298}([M+H]^+(H_2O)_n)_{sc} - H^\circ_{298}([M+H]^+(H_2O)_{n-1})_{sc} - H^\circ_{298}(H_2O) \quad (2_n)$$

$$\Delta S^\circ_{n,sc} = S^\circ_{298}([M+H]^+(H_2O)_n)_{sc} - S^\circ_{298}([M+H]^+(H_2O)_{n-1})_{sc} - S^\circ_{298}(H_2O) \quad (3_n)$$

$$\Delta G^\circ_{n,sc} = G^\circ_{298}([M+H]^+(H_2O)_n)_{sc} - G^\circ_{298}([M+H]^+(H_2O)_{n-1})_{sc} - G^\circ_{298}(H_2O) \quad (4_n)$$

where  $H^\circ_{298}$ ,  $S^\circ_{298}$  and  $G^\circ_{298}$  represent standard enthalpy, entropy and standard Gibbs free energy at 298 K, respectively. These “same conformation” data allow us to estimate the strength of a given non-covalent interaction for a given conformer of  $G_n$ .

In fact, intrinsic hydration thermochemical data of a given species should consider one mol of reactant and products in thermal equilibrium at 298 K. The corresponding populations of N conformers should define a Boltzmann distribution where the molar fractions  $x_i$  of each conformer at temperature 298 K is given by equation (5).

$$x_i = \exp(-G^\circ_{298,i}/RT) / \sum_1^N \exp(-G^\circ_{298,i}/RT) \quad (5)$$

where  $G^\circ_{298,i}$  are the individual Gibbs free energies of each conformers. Averaged enthalpy value of one mole of mixture can be determined using the most stables structures of the same generation according to equation (6)

$$H^\circ_{298,average} = \sum_1^N x_i (H^\circ_{298})_i \quad (6)$$

Thus leading to:

$$\Delta H^\circ_{n,average} = H^\circ_{298}([M+H]^+(H_2O)_n)_{average} - H^\circ_{298}([M+H]^+(H_2O)_{n-1})_{average} - H^\circ_{298}(H_2O) \quad (7)$$

Similarly, “average” definition can be introduced for entropy,  $\Delta S^\circ_{n,average}$  and Gibbs free energy,  $\Delta G^\circ_{n,average}$ . However, it should be noted that the averaged entropy of a collection of conformers should include the entropy of mixing and be consequently estimated via equation (8):

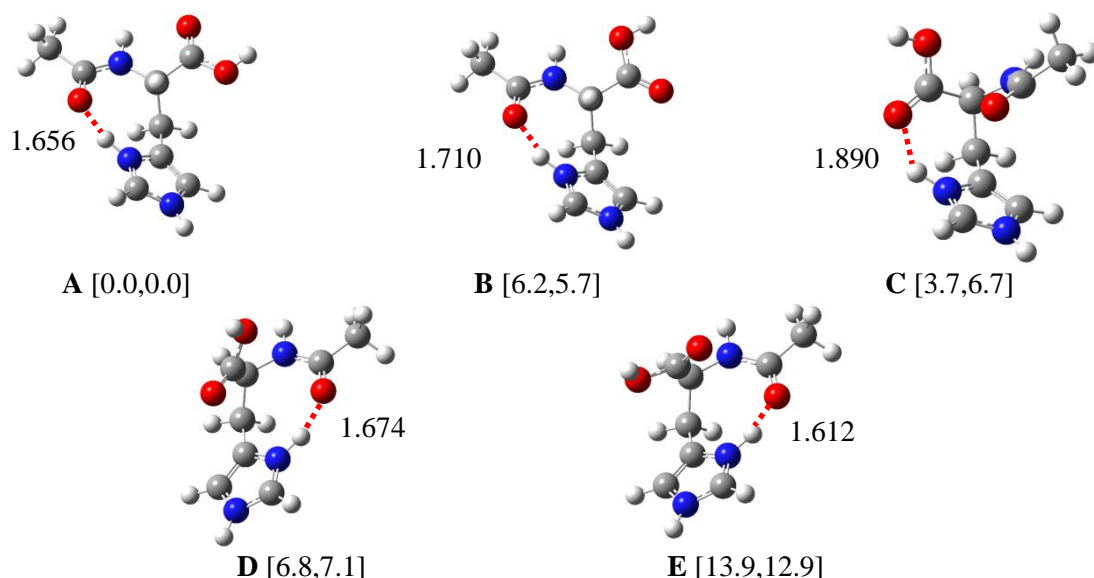
$$S^\circ_{T,average} = \sum_1^N x_i (S^\circ_T)_i - R \sum_1^N x_i \ln x_i \quad (8)$$

where the second component corresponds to the mixing contribution. This should be also considered during the  $\Delta G^\circ_{n,average}$  calculation.

## RESULTS AND DISCUSSION

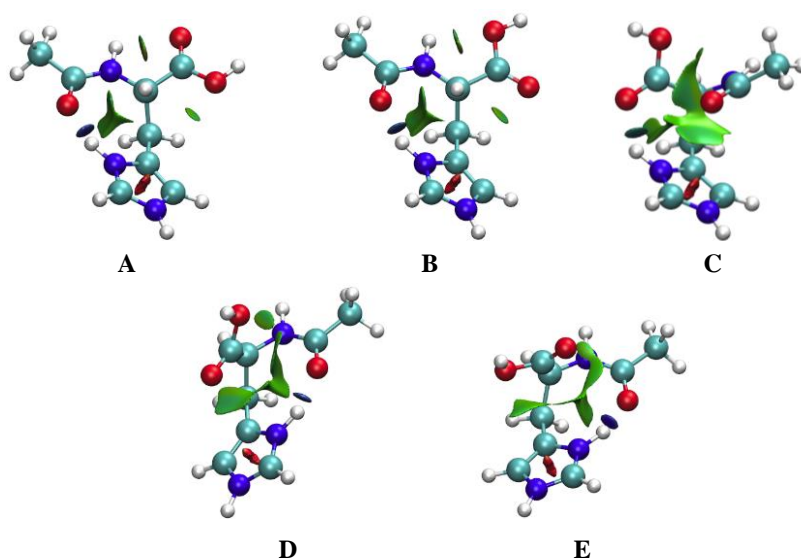
In the following section we will present the main results of our exploration of the conformational landscapes.

**Non-hydrated N-Acetylhistidine,  $[NAH+H]^+$ , population of conformers  $G_0$ .** This system includes (1) two cationic hydrogen bond donor sites,  $N^{\delta}H$  (site a, Figure 1) and  $N^{\epsilon}H$  (site b) ; (2) two neutral hydrogen bond donor sites, that is the amide NH (site c) and the carboxylic OH (site d) groups; (3) three hydrogen bond acceptors sites (sites e, f and g). Therefore, intramolecular H-bond seems possible, in particular with charged sites a or b. However, due to geometric reasons, site b is unable to form any intramolecular H-bond. On the opposite, all most stable conformers in the 0-15 kJ mol<sup>-1</sup> range in Gibbs free energy include an intramolecular H-bond through site a (Figure 2).



**Figure 2.** Optimized geometries (M06/6-311++G(d,p)) of the five most stable conformers of  $[\text{NAH}+\text{H}]^+$  (in brackets, relative  $H^\circ_{298}$  and  $G^\circ_{298}$  in  $\text{kJ mol}^{-1}$ ).

Conformers **A**, **B**, **D** and **E** present a  $\text{N}^\delta\text{H}^\delta\cdots\text{OCNH}$  hydrogen bond, whereas **C** presents a  $\text{N}^\delta\text{H}^\delta\cdots\text{OCOH}$  interaction. The  $\text{N}^\delta\text{H}^\delta\cdots\text{OCNH}$  hydrogen interaction is particularly efficient because it involves strong acidic hydrogen (site a) and the most basic oxygen (site f). This is attested to by the short  $\text{O}\cdots\text{H}$  distances (1.612-1.710 Å). Structures **A** and **B** differ only by the orientation of the carboxyl acid moiety, and the same difference in the orientation of the carboxyl group is observed for **D** and **E**. The relative Gibbs free energies  $G^\circ_{298}$ , which give the stability order of these structures, follow the same order as the relative enthalpies  $H^\circ_{298}$ , excepted between structures **B** and **C**. For the latter, we note that the oxygen of the amide group (site f) is oriented towards the imidazolium ring, inducing a weak stabilizing interaction as reveals by the Non-Covalent Interaction (NCI) analysis (Figure 3). So **C** is more rigid than **B**, as confirmed by the entropy which is lower for **C** ( $S^\circ_{298}(\text{B}) = 487.8 \text{ J K}^{-1} \text{ mol}^{-1}$  and  $S^\circ_{298}(\text{C}) = 476.4 \text{ J K}^{-1} \text{ mol}^{-1}$ ), explaining the stabilization order inversion. The NCI analysis reveals also the presence of minor interactions implying sites c, e and/or g, among which  $\text{N}_{\text{amide}}\text{H}^\delta\cdots\text{OC}_{\text{COOH}}$  hydrogen bond (**A** and **B**) and  $\text{CO}_{\text{COOH}}\cdots\pi$  interaction (**D** and **E**), the latter being less stabilizing. Using the computed  $G^\circ_{298}$  values, a ratio of **A**, **B**, **C**, **D** and **E** equal to 81/8/6/5/0 % is estimated. We can therefore assume that the most stable structures of  $\text{G}_1$  should be of type **A**, for which the first-shell hydration molecules can interact with H-bond donor (b, c and d) or acceptor (e and g) sites.



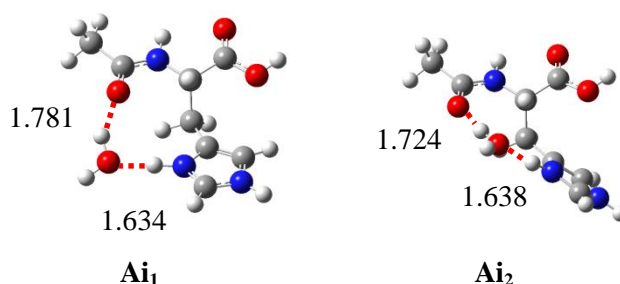
**Figure 3.** The nature of non-covalent interactions within the structures **A**, **B**, **C**, **D** and **E**.

**Monohydrated N-Acetylhistidine,  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})$ , population of conformers  $\text{G}_1$ .** Based on the hierarchical approach, we have found 8 structures in the 0-15  $\text{kJ mol}^{-1}$  range both at the M06/6-31+G(d,p) and M06/6-311++G(d,p) levels (Table 1). As expected due to the cationic nature of site b, formation of a H-bond between the water molecule and site b of the lowest energy structure **A** leads to the  $\text{G}_1$  global minima, noted **Ab**. Other minima include children from  $\text{G}_0$  structure with the same H-bond and children of **A** with H-bond to site d. Three special structures (**Ai<sub>1</sub>**, **Ai<sub>2</sub>** and **Bi<sub>2</sub>**) are identified at 6.2, 11.1 and 12.4  $\text{kJ mol}^{-1}$  relative Gibbs free energy (Figure 4). In these cases, the water molecule is inserted into the intramolecular H-bond of **A** or **B**, inducing noticeable structural modifications. The dihedral angle  $\text{H-C}_\alpha\text{-C}_\beta\text{-C}_\gamma$  for **A**, **Ai<sub>1</sub>** and **Ai<sub>2</sub>** are respectively equal to  $5.6^\circ$ ,  $-24.0^\circ$  and  $-48.4^\circ$ . Regarding the low relative Gibbs free energy of **Ai<sub>1</sub>** as well as the similarities between the relative Gibbs free energies of **A**, **B** and **C** (0.0, 5.7 and 6.7  $\text{kJ mol}^{-1}$ ) and **Ab**, **Bb** and **Cb** (0.0, 5.7 and 8.3  $\text{kJ mol}^{-1}$ ), we suspected that other conformers such as **Bi<sub>1</sub>** or **Db** should be located in the 0-15  $\text{kJ mol}^{-1}$  Gibbs free energy window. To verify our hypothesis and control the population of  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})$ , the Darwinian tree starting from  $\text{G}_0$  has been constructed considering all H-bond acceptor and donor represented in Figure 1 as well as the two insertion sites  $i_1$  and  $i_2$ . Three new structures (**Bi<sub>1</sub>**, **Db** and **Eb**) in the 0-15  $\text{kJ mol}^{-1}$  range at the M06/6-311++G(d,p) level have been obtained. Re-optimization of these structures at the AMOEBA and M06/6-31+G(d,p) allow to understand why they have not been identified in the hierarchical approach (Table 1). The relative Gibbs free energies for **Db** and **Eb** are also close to those of **D** and **E** (7.2 and 13.9  $\text{kJ mol}^{-1}$ ).

**Table 1.**  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})$  relative energies, enthalpies, entropies and Gibbs free energies computed at M06/6-311++G(d,p) in  $\text{kJ mol}^{-1}$ , except  $S^\circ_{298}$  in  $\text{J. mol}^{-1} \text{K}^{-1}$ .

Structures	$\Delta E^\circ_{298}$	$\Delta H^\circ_{298}$	$S^\circ_{298}$	$\Delta G^\circ_{298}$
<b>Ab</b>	0.0	0.0	564.5	0.0
<b>Ad</b>	2.4	2.0	553.7	5.2
<b>Ai<sub>1</sub></b>	10.0	9.8	576.6	6.2
<b>Ai<sub>2</sub></b>	5.8	6.3	548.3	11.1
<b>Bb</b>	5.7	6.0	565.2	5.7
<b>Bi<sub>1</sub><sup>a</sup></b>	6.6	7.3	553.5	10.6
<b>Bd</b>	8.2	8.3	553.3	11.6
<b>Bi<sub>2</sub></b>	5.8	6.7	545.6	12.4
<b>Cb</b>	3.8	4.8	552.6	8.3
<b>Db<sup>b</sup></b>	8.6	8.6	569.4	7.2
<b>Eb<sup>b</sup></b>	13.7	14.0	564.7	13.9

<sup>a</sup> Missing minima in the AMOEBA potential energy surface. <sup>b</sup> The relative energy of the structure is larger than to 20  $\text{kJ mol}^{-1}$  with AMOEBA.

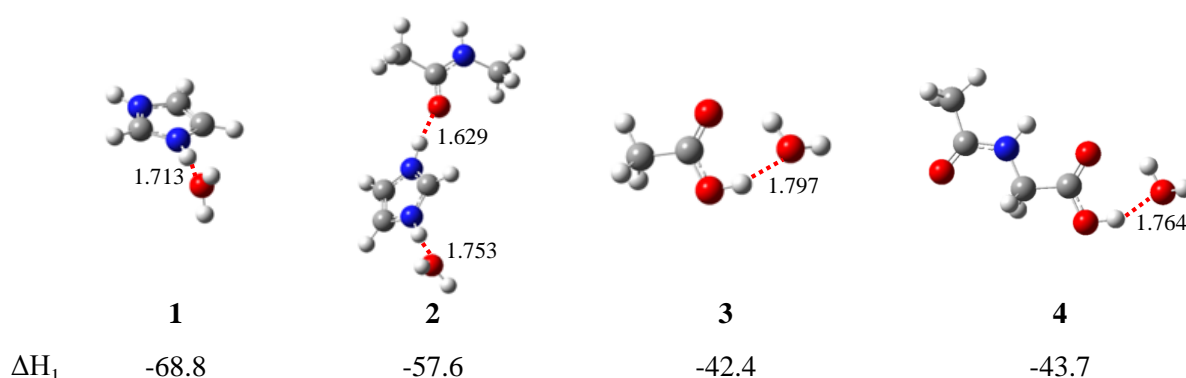


**Figure 4.** Optimized geometries (M06/6-311++G(d,p)) of structures **Ai<sub>1</sub>** and **Ai<sub>2</sub>**.

As for generation  $\text{G}_0$ , differences are observed in the stability order of the conformers in terms of  $H^\circ_{298}$  and  $G^\circ_{298}$ . These trends could be explained by the entropic contribution which is not negligible especially when water molecule is located in the  $i_1$  or  $i_2$  sites.

Theoretical hydration enthalpies, entropies and Gibbs free energies of  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})$  are summarized in Table 2. It may be underlined that all  $\Delta H^\circ_{n,\text{average}}$ ,  $\Delta S^\circ_{n,\text{average}}$  and  $\Delta G^\circ_{n,\text{average}}$  values reported in Table 2 are defined with respect to the most stables conformers in the  $G^\circ_{298}$  scales. A threshold of 3 RT was used to define the population of conformers used in eq. 5-7, all structures below

this limit represent a negligible percentage  $x$ , (eq. 5) and have therefore been excluded from the calculation. As expected,  $\Delta H_{1,sc}^\circ$  for **Ab**, **Bb** and **Db** are almost similar, and the same is true for their  $\Delta G_{1,sc}^\circ$  values, indicating that the strength of the H-bond through site b is independent of the conformation. Averaging values for **Ab**, **Bb** and **Db** lead to  $\Delta H_{1,average}^\circ(\text{site b}) = -57.8 \text{ kJ mol}^{-1}$  and  $\Delta G_{1,average}^\circ(\text{site b}) = -25.4 \text{ kJ mol}^{-1}$ . To investigate the effect of the intramolecular H-bond on these values, we evaluate the strength of the imidazolium $\cdots$ OH<sub>2</sub> H-bond for small models **1** and **2** (Figure 5). The mono-hydration enthalpies calculated for **1** and **2** are  $-68.8$  and  $-57.6 \text{ kJ mol}^{-1}$ , respectively. This lead to the conclusions that the amide $\cdots$ imidazolium H-bond reduces the strength of the imidazolium $\cdots$ OH<sub>2</sub> H-bond by more than  $10 \text{ kJ mol}^{-1}$  and that the C $_{\alpha}$ -COOH group, not presents in model **2**, does not influence the coordination of the water molecule. Surprisingly, the sc hydration values for conformer **Ad** are similar to the one computed for the site b, even if the H-bond implies only neutral groups. Comparison with models **3** and **4** reveals the long range influence of the positive charge brought by the imidazolium group on the strength of the intermolecular H-bond in **Ad**. Furthermore, this indicates that experimental hydration enthalpies values obtained for hydrated models systems including only one imidazolium group, do not allow to reveal the location of the added water molecule.



**Figure 5.** Optimized geometries (M06/6-311++G(d,p)) of models. Hydration enthalpies are in  $\text{kJ mol}^{-1}$ .

**Table 2.** Computed hydration enthalpies and entropies of  $[\text{NAH}+\text{H}]^+$  obtained from theory (M06/6-311++G(d,p)).

$G_n$	structure	ratio	$\Delta H_n^\circ (\text{kJ mol}^{-1})$	$\Delta G_n^\circ (\text{kJ mol}^{-1})$	$\Delta S_n^\circ (\text{J K}^{-1} \text{mol}^{-1})$
$G_1$	<b>Ab</b>	0.73	-58.4 <sup>a</sup>	-25.4 <sup>a</sup>	-110.6 <sup>a</sup>
	<b>Ad</b>	0.09	-56.4 <sup>a</sup>	-20.2 <sup>a</sup>	-121.4 <sup>a</sup>
	<b>Bb</b>	0.08	-58.6 <sup>a</sup>	-25.4 <sup>a</sup>	-111.3 <sup>a</sup>
	<b>Ai<sub>1</sub></b>	0.06	-48.5 <sup>a</sup>	-19.2 <sup>a</sup>	-98.5 <sup>a</sup>
	<b>Db</b>	0.04	-56.5 <sup>a</sup>	-25.3 <sup>a</sup>	-104.7 <sup>a</sup>
	Total	1.00			
	<b>average</b>		<b>-57.8<sup>b</sup></b>	<b>-25.7 (-25.0)<sup>b</sup></b>	<b>-107.9 (-110.1)<sup>b</sup></b>
$G_2$	<b>Abd</b>	0.68	-55.3 <sup>a</sup>	-20.2 <sup>a</sup>	-117.6 <sup>a</sup>
	<b>Abw<sub>2</sub></b>	0.09	-45.4 <sup>a</sup>	-15.2 <sup>a</sup>	-101.4 <sup>a</sup>
	<b>Abi<sub>2</sub></b>	0.09	-51.2 <sup>a</sup>	-15.1 <sup>a</sup>	-121.2 <sup>a</sup>
	<b>Bbi<sub>2</sub></b>	0.05	-56.3 <sup>a</sup>	-19.5 <sup>a</sup>	-123.8 <sup>a</sup>
	<b>Bbd</b>	0.05	-54.5 <sup>a</sup>	-19.4 <sup>a</sup>	-117.7 <sup>a</sup>
	<b>Abw<sub>1</sub></b>	0.04	-45.2 <sup>a</sup>	-13.2 <sup>a</sup>	-107.5 <sup>a</sup>
	Total	1.00			
	<b>average</b>		<b>-54.6<sup>b</sup></b>	<b>-20.4 (-20.0)<sup>b</sup></b>	<b>-114.8 (-116.3)<sup>b</sup></b>
$G_3$	<b>Abdi<sub>2</sub></b>	0.29	-51.8 <sup>a</sup>	-13.9 <sup>a</sup>	-127.3 <sup>a</sup>
	<b>Abdw<sub>1</sub></b>	0.26	-44.2 <sup>a</sup>	-13.6 <sup>a</sup>	-103.0 <sup>a</sup>
	<b>Abdw<sub>2</sub></b>	0.15	-44.6 <sup>a</sup>	-12.1 <sup>a</sup>	-108.9 <sup>a</sup>
	<b>Bbdw<sub>2</sub></b>	0.06	-44.7 <sup>a</sup>	-16.6 <sup>a</sup>	-94.0 <sup>a</sup>
	<b>Abdh</b>	0.05	-31.9 <sup>a</sup>	-9.4 <sup>a</sup>	-75.4 <sup>a</sup>

<b>Bbdi<sub>1</sub></b>	0.05	-50.8 <sup>a</sup>	-15.8 <sup>a</sup>	-117.6 <sup>a</sup>
<b>Bbi<sub>2</sub>d</b>	0.04	-51.6 <sup>a</sup>	-15.5 <sup>a</sup>	-121.2 <sup>a</sup>
<b>Bbdw<sub>1</sub></b>	0.04	-44.6 <sup>a</sup>	-15.6 <sup>a</sup>	-97.5 <sup>a</sup>
<b>Abi<sub>2</sub>w<sub>2</sub></b>	0.02	-50.7 <sup>a</sup>	-12.3 <sup>a</sup>	-128.6 <sup>a</sup>
<b>Bbi<sub>2</sub>w<sub>2</sub></b>	0.02	-43.6 <sup>a</sup>	-13.8 <sup>a</sup>	-100.3 <sup>a</sup>
<b>Ab{de}<sub>1</sub></b>	0.02	-46.8 <sup>a</sup>	-6.9 <sup>a</sup>	-133.8 <sup>a</sup>
total	1.00			
<b>average</b>		<b>-47.4<sup>b</sup></b>	<b>-16.0 (-13.9)<sup>b</sup></b>	<b>-105.6 (-112.6)<sup>b</sup></b>

<sup>a</sup> “same conformation” values. <sup>b</sup> “average” values calculated over the 298 K distribution of conformers. Data without entropy of mixing are indicated in parentheses.

**Dihydrated N-Acetylhistidine, [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>2</sub>, population of conformers G<sub>2</sub>.** 14 minimas in the 0-15 kJ mol<sup>-1</sup> range have been identified during the hierarchical process at the M06/6-31+G(d,p). Among them, 10 are preserved at the M06/6-311++G(d,p) (Table 4). 6 are related to conformer **A**, and 3 to conformer **B**. A new conformation of protonated N $\alpha$ -acetylhistidine, which do not include intramolecular H-bond, is identified as conformer **Fbi<sub>1</sub>** (Figure 6). Complementary calculations confirm that structures **F** (**G<sub>0</sub>**), **Fb** and **Fi<sub>1</sub>** (**G<sub>1</sub>**) are high energy (> 15 kJ mol<sup>-1</sup>) conformer in their respective generation. Complementing the Darwinian tree, which can be built from the previous structures, allow to obtain 11 new minima in the 0-15 kJ mol<sup>-1</sup> range whose the lowest in relative Gibbs free energy is **Abw<sub>1</sub>** located at 5.0 kJ mol<sup>-1</sup>. The water molecule can form H-bond to previously observed sites (b, d, c, i<sub>1</sub> and i<sub>2</sub>) or at unobserved sites in previous generation. Site j in **Cbj** denotes the insertion of the water molecule into the intramolecular bond between sites a and e observed for **C**. Site h in **Abh** indicates formation of a C<sub>imidazolium</sub>H<sup>+</sup>OH<sub>2</sub> H-bond. Last, the second water molecule can be located in the second solvation shell through formation of a H-bond with the first water molecule. Different relative orientation of these water molecules induces several conformers noted w<sub>1</sub>, w<sub>2</sub>, ....

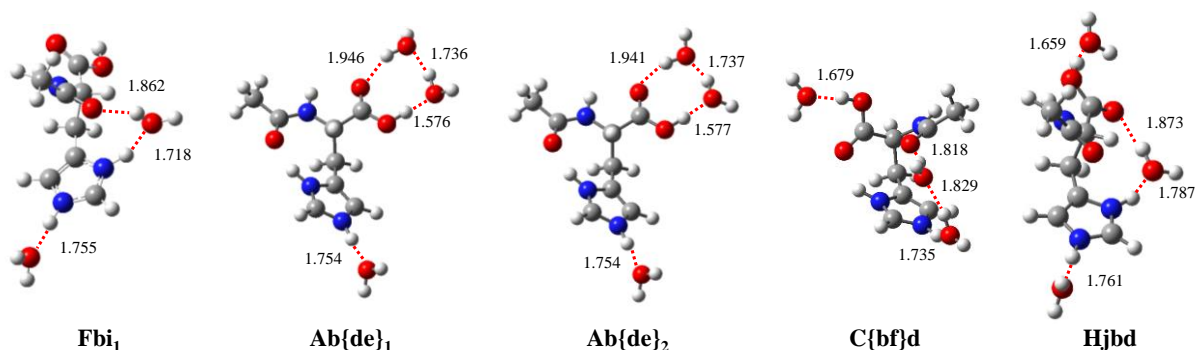
**Table 3.** [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>2</sub> and [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>3</sub> relative energies, enthalpies and Gibbs free energies computed at M06/6-311++G(d,p) in kJ mol<sup>-1</sup>, except S<sup>o</sup><sub>298</sub> in J mol<sup>-1</sup> K<sup>-1</sup>.

Structures	$\Delta E^\circ_{298}$	$\Delta H^\circ_{298}$	$S^\circ_{298}$	$\Delta G^\circ_{298}$	Structures	$\Delta E^\circ_{298}$	$\Delta H^\circ_{298}$	$S^\circ_{298}$	$\Delta G^\circ_{298}$
[NAH+H] <sup>+</sup> (H <sub>2</sub> O) <sub>2</sub>					[NAH+H] <sup>+</sup> (H <sub>2</sub> O) <sub>3</sub>				
<b>Abd</b>	0.0	0.0	635.6	0.0	<b>Abdi<sub>2</sub></b>	0.3	0.0	696.9	0.0
<b>Abw<sub>2</sub><sup>a</sup></b>	9.5	9.8	651.7	5.0	<b>Abdw<sub>1</sub></b>	7.8	7.6	721.2	0.3
<b>Abi<sub>2</sub></b>	3.0	4.1	632.0	5.1	<b>Abdw<sub>2</sub></b>	7.4	7.2	715.3	1.8
<b>Abw<sub>1</sub></b>	9.6	10.1	645.7	7.0	<b>Abdh<sup>b</sup></b>	20.4	20.0	748.8	4.5
<b>Abc</b>	14.2	14.4	644.6	11.7	<b>Abi<sub>2</sub>w<sub>2</sub></b>	10.0	11.0	711.8	6.6
<b>Abh<sup>a/b</sup></b>	21.8	21.7	668.6	11.9	<b>Ab{de}<sub>1</sub></b>	5.4	5.0	690.4	6.9
<b>Adi<sub>2</sub></b>	3.7	4.6	606.8	13.1	<b>Abi<sub>2</sub>w<sub>1</sub><sup>b</sup></b>	10.3	11.0	707.5	7.8
<b>Abw<sub>3</sub></b>	10.8	12.3	630.9	13.7	<b>Ab{de}<sub>2</sub></b>	5.9	5.5	688.0	8.2
<b>Bbi<sub>2</sub></b>	4.3	4.9	630.1	6.5	<b>Abw<sub>1</sub>w<sub>3</sub><sup>b</sup></b>	22.9	23.1	745.0	8.8
<b>Bbd</b>	6.4	6.8	636.2	6.6	<b>Abdw<sub>3</sub><sup>a</sup></b>	9.0	9.7	699.1	9.0
<b>Bbw<sub>1</sub></b>	15.5	15.9	656.5	9.6	<b>Abw<sub>2</sub>c<sup>b</sup></b>	21.7	20.9	731.7	10.6
<b>Bbw<sub>2</sub><sup>a</sup></b>	15.4	15.9	653.5	10.6	<b>Abw<sub>2</sub>w<sub>4</sub><sup>a/b</sup></b>	22.5	23.1	737.3	11.1
<b>Bdi<sub>2</sub><sup>c</sup></b>	7.2	7.6	624.4	10.9	<b>Abi<sub>2</sub>c<sup>c</sup></b>	15.9	15.7	712.3	11.2
<b>Bi<sub>1</sub>d<sup>a/d</sup></b>	8.9	9.4	625.4	12.5	<b>Abdc<sup>b</sup></b>	15.1	14.3	706.2	11.5
<b>Cbw<sub>1</sub><sup>b</sup></b>	14.7	15.5	659.3	8.5	<b>Abi<sub>2</sub>h<sup>a/b</sup></b>	23.2	23.4	728.5	14.0
<b>Cbj<sup>b/c</sup></b>	13.9	14.3	644.7	11.5	<b>Abw<sub>1</sub>w<sub>4</sub><sup>b</sup></b>	23.1	23.2	725.1	14.8
<b>Cbd<sup>b</sup></b>	8.7	9.5	626.1	12.4	<b>Bbdw<sub>2</sub></b>	14.3	13.9	730.8	3.8
<b>Cbw<sub>2</sub><sup>b/c</sup></b>	13.9	14.9	641.9	13.0	<b>Bbdi<sub>1</sub><sup>d</sup></b>	7.4	7.7	707.2	4.7
<b>Dbw<sub>1</sub><sup>b/c</sup></b>	19.5	20.4	658.3	13.7	<b>Bbi<sub>2</sub>d</b>	4.6	5.1	697.6	4.9
<b>Dbd<sup>b/c</sup></b>	13.7	14.3	633.6	14.9	<b>Bbdw<sub>1</sub></b>	14.4	13.9	727.4	4.9
<b>Fbi<sub>1</sub></b>	9.8	12.0	629.3	13.9	<b>Bbi<sub>2</sub>w<sub>2</sub></b>	12.3	13.1	718.5	6.6
					<b>Bbi<sub>2</sub>w<sub>1</sub></b>	12.7	13.3	714.6	8.0
					<b>Bbdh<sup>b/c</sup></b>	27.2	26.2	753.9	9.2
					<b>Bb{de}<sub>2</sub><sup>b</sup></b>	11.7	11.4	690.6	13.3
					<b>Bi<sub>1</sub>dx<sup>d/c</sup></b>	21.4	21.1	721.9	13.6
					<b>C{bf}d<sup>c</sup></b>	0.0	1.5	654.3	14.2
					<b>Ebi<sub>1</sub>d</b>	5.4	6.7	680.6	11.6
					<b>Fbi<sub>1</sub>d</b>	9.7	10.5	694.6	11.2
					<b>Fbi<sub>1</sub>w<sub>1</sub></b>	17.6	18.5	715.1	13.1



<b>Fbi<sub>1</sub>w<sub>2</sub></b>	17.2	18.2	709.9	14.4
<b>Hjbd</b>	13.5	14.6	696.1	14.8

<sup>a</sup> Missing minima in the potential energy surface of M06/6-31+G(d,p). <sup>b</sup> The relative energy of the structure is larger than 20 kJ mol<sup>-1</sup> with AMOEBA. <sup>c</sup> The relative Gibbs free energy of the structure is superior to 15 kJ mol<sup>-1</sup> with M06/6-31+G(d,p). <sup>d</sup> Missing minima in the potential energy surface of AMOEBA.



**Figure 6.** Optimized geometries (M06/6-311++G(d,p)) of structures **Fbi<sub>1</sub>**, **Ab{de}<sub>1</sub>**, **Ab{de}<sub>2</sub>**, **C{bf}d** and **Hjbd**.

Assuming that the most stable conformer of **G<sub>n-1</sub>** gives the best conformation of **G<sub>n</sub>**, and with d the second preferential site of **A**, it is not surprising that **Abd** would be the most stable conformer in **G<sub>2</sub>**. Other low energy structures always include a water molecule bound to site b and the second one bound to either i<sub>2</sub>, d or w. As observed previously the relative energy of **Bbd** (6.6 kJ mol<sup>-1</sup>) is similar to the relative energy of **B** in **G<sub>0</sub>** and **Bb** in **G<sub>1</sub>** indicating similar H-bond strength in **Abd** and **Bbd**. On the opposite, the relative energies of **Cbd** (12.4 kJ mol<sup>-1</sup>) and **Dbd** (14.9 kJ mol<sup>-1</sup>) are significantly larger than ones observed for **C** and **D** or for **Cb** and **Db**. This reveals that the strength of the H-bond formed by the water molecules to the imidazolium and the COOH groups are lowered, presumably due to the interaction between these two groups in **C** and **D** (Figure 3). The structure **Abd** is the main conformer with a ratio to 68 %; so  $\Delta X^{\circ}_{2,sc}$  (X=H,G and S) are close to  $\Delta X^{\circ}_{2,average}$  with a deviation of -0.7, 0.2 kJ mol<sup>-1</sup> and -2.8 J K<sup>-1</sup> mol<sup>-1</sup>, respectively (Table 2). Furthermore,  $\Delta H^{\circ}_{2,average}$  is less favorable by ~ 10 kJ mol<sup>-1</sup> for site w than for site d. The H-bond between the water molecule in the first and second solvation shell in **Abw<sub>2</sub>** is weaker in enthalpy ( $\Delta H^{\circ}_{2,sc} = -45.4$  kJ mol<sup>-1</sup>) than the interactions formed after insertion into the intramolecular H-bond in **Abi<sub>2</sub>** ( $\Delta H^{\circ}_{2,sc} = -51.2$  kJ mol<sup>-1</sup>), but these two conformers are in the same abundance (9 %) due to entropic terms which favour **Abw<sub>2</sub>**. Based on these conformers, it can be postulated that structures **Abdi<sub>2</sub>** or **Abdw** should be among the most stable conformers in **G<sub>3</sub>**.

**Tri-hydrated N-Acetylhistidine, [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>3</sub>, population of conformers **G<sub>3</sub>**.** The hierarchical method leads to 19 minima in the 0-15 kJ mol<sup>-1</sup> range in M06/6-31+G(d,p) and 16 are conserved at the M06/6-311++G(d,p) level (see Table 3). A new structure **H** is identified (**Hjbd**), its backbone differs from **F** only by the orientation of the carboxyl acid moiety (Figure 6). As for **G<sub>1</sub>** and **G<sub>2</sub>**, the Darwinian tree strategy allow to complete the population in the 0-15 kJ mol<sup>-1</sup> range by addition of 15 structures among which the most stable conformer, **Abdh**, is at  $\Delta G^{\circ}_{298} = 4.5$  kJ mol<sup>-1</sup>, and two conformers (**Abdh** and **Bbdi<sub>1</sub>**) include the threshold of 3 RT. Not surprisingly, some structures have water molecule in the third solvation shell (noted w') or in the second solvation shell of site d (noted x). Furthermore, we observed that two water molecules can bridge sites d and e (noted {de}) or sites b and f (noted {bf}) (Figure 6). As expected, **Abdi<sub>2</sub>** and **Abdw** are the most stable structures and represent together ~ 70 % of the Boltzmann population at 298 K. More precisely the larger ratios are obtained for **Abdi<sub>2</sub>**, **Abdw<sub>1</sub>** and **Abdw<sub>2</sub>** with 29, 26 and 15 % of the population. It is interesting to note that the lowest Gibbs free energy conformer does not include anymore an intramolecular H-bond, as the third water molecule is inserted into it. Considering the Boltzmann population at 298 K, a water molecule is located in the second solvation shell for 55 % of structures while the breakage of the intramolecular H-bond or the bridging is present for 36 % of population. The sc values for i<sub>2</sub> and w sites are similar to the values obtained in previous generation. In average, the hydration values are



significantly reduced compared to **G**<sub>2</sub>, mainly due to non-negligible ratio of **Abdw** conformers which have lower enthalpy values.

The entropy of mixing affects further  $\Delta S_{3,\text{average}}^\circ$  and  $\Delta G_{3,\text{average}}^\circ$  due to the increase of the number of water molecules and structures in the population. For **G**<sub>0</sub>, **G**<sub>1</sub>, **G**<sub>2</sub> and **G**<sub>3</sub>,  $S_{\text{mix}}$  is equal to 5.7, 7.8, 9.3 and 16.3 J K<sup>-1</sup> mol<sup>-1</sup> respectively. As evidenced by the decrease of the  $\Delta H_{n,\text{average}}^\circ$  and  $\Delta G_{n,\text{average}}^\circ$ , the subsequent interaction of water molecule is progressively weaker as already mentioned in the literature for systems including ammonium groups.<sup>4,7,8,10,11,55,56</sup> From this study of **G**<sub>0</sub> to **G**<sub>3</sub>, we note that the successive addition of water molecules to the protonated N $\alpha$ -acetylhistidine mainly preserved the backbone of the lowest energy structures, the largest deviation being observed upon the insertion (*i*<sub>2</sub>) into the intramolecular H-bond. Conformer of type **A** remains the most stable structure of all generations following the enthalpy and Gibbs free energy scales. It should however be noted that in the energy scale, conformer **C{bf}d** is the most stable in **G**<sub>3</sub>. Last, to be effective and to correspond to the most stable conformer, the breakage of the intramolecular H-bond requires three water molecules.

## CONCLUSION

Using sequentially the hierarchical and the Darwinian tree methods, we have exhaustively identified the conformer population of [NAH+H]<sup>+</sup> in the 0-15 kJ mol<sup>-1</sup> range in Gibbs free energy as well as in the children generations obtained by addition of one to three molecules of water. Among the 68 structures obtained for the four generations, approximately 60 % have been obtained from the hierarchical approach and the remaining 40 % result from the Darwinian tree method. All the lowest energy structures have been obtained after the hierarchical method step, but four (nine) structures with relative Gibbs free energy lower than 7.5 kJ mol<sup>-1</sup> (10.0 kJ mol<sup>-1</sup>) have been reached only after the Darwinian tree step. That is to say 18 % of structures used to calculate the thermochemical quantities, in applying a threshold of 3 RT, results only from the Darwinian tree approach. The entropic contribution plays a noticeable influence given the relative stabilities for enthalpy and Gibbs free energy. This effect increases with the number of the water molecule as shown by the mean absolute deviations between  $H_{298}^\circ$  and  $G_{298}^\circ$  which are equal to 0.96, 2.65, 3.57 and 6.14 for **G**<sub>0</sub>, **G**<sub>1</sub>, **G**<sub>2</sub> and **G**<sub>3</sub>, respectively. So, to have a complete population in the 0.0 - 7.5 kJ mol<sup>-1</sup> range in Gibbs free energy to compute accurately the thermochemical data, it is necessary to select a larger range in energy during the potential energy surface exploration. The relative stability of each n-hydrated structure for each generation reveals the preference for different available hydration sites. The first water molecule is located on the imidazolium NH group, and the second on the COOH group. Two sites are in competition for the third hydration: the breakage of the intramolecular H-bond and the interaction with the water molecule which already interact with the cationic group. The backbone of [NAH+H]<sup>+</sup> is generally well preserved, except in the case of the breakage of the intramolecular H-bond or the bridging of neighboring sites, where the imidazolium ring is reoriented. Considering the ratio of each conformer, we concluded that considering only one conformer is not enough to interpret the average thermochemical quantities for **G**<sub>3</sub>, but works if one conformer is clearly dominant as in **G**<sub>1</sub> and **G**<sub>2</sub>. We postulate that the **G**<sub>3</sub> case is more general for flexible systems or those including more water molecules, indicating that experimental values could not be related to a single conformation. Examination of the strength of the H-bond, either through small models or by calculating  $\sigma$  values indicates noticeable reciprocal influence between coulombic, intra- and intermolecular H-bonds non-covalent interactions. The decreasing values of  $\Delta H_n^\circ$  and  $\Delta G_n^\circ$  when increasing *n* values confirm that the binding of a water molecule to a cationic species decreases its ability to form additional intermolecular H-bond.

## SUPPORTING INFORMATION

Complete author lists for refs (41) and (51). [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>*n*</sub> relative energies, enthalpies and Gibbs free energies computed at M06/6-31+G(d,p) and M06/6-311++G(d,p) in kJ mol<sup>-1</sup>. Cartesian coordinates and molecular drawing for all conformers in the 0-15 kJ mol<sup>-1</sup> range of **G**<sub>0</sub>-**G**<sub>3</sub>. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ACKNOWLEDGMENTS

This work was performed using HPC resources from GENCI-CINES (grant x2013087105). Dr. E. Rezabal is acknowledged for her help to NCI calculations.

## REFERENCES

- (1) Pratt, L. R.; Pohorille, A. *Chem. Rev.* **2002**, *102*, 2671-2692.
- (2) Levy, Y.; Onuchic, J. N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389-415.
- (3) Blades, A. T.; Klassen, J. S.; Kebarle, P. J. *Am. Chem. Soc.* **1995**, *117*, 10563-10571.
- (4) Gilligan, J. J.; Lampe, F. W.; Nguyen, V. Q.; Vieira, N. E.; Yergey, A. L. *J. Phys. Chem. A* **2003**, *107*, 3687-3691.
- (5) Meot-Ner, M. *Chem. Rev.* **2005**, *105*, 213-284.
- (6) Meot-Ner, M. *Chem. Rev.* **2012**, *112*, RP22-RP103.
- (7) Wincel, H. *Int. J. Mass Spectrom.* **2006**, *251*, 23-31.
- (8) Wincel H. *Chem. Phys. Lett.* **2007**, *439*, 157-161.
- (9) Liu, D.; Wyttenbach, T.; Barran, P. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2003**, *125*, 8458-8464.
- (10) Wyttenbach, T.; Liu, D.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 221-232.
- (11) Wyttenbach, T.; Bowers, M. T. *Chem. Phys. Lett.* **2009**, *480*, 1-16.
- (12) Gao, B.; Wyttenbach, T.; Bowers, M. T. *J. Am. Chem. Soc.* **2009**, *131*, 4695-4701.
- (13) Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2002**, *124*, 11148-11158.
- (14) Lemoff, A. S.; Bush, M. F.; Wu, C.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 13576-13584.
- (15) Ye, S. J.; Moision, R. M.; Armentrout, P. B.; *Int. J. Mass Spectrom.* **2006**, *253*, 288-304.
- (16) Prell, J. S.; Chang, T. M.; O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 7811-7819.
- (17) Scuderi, D.; Bakker, J. M.; Durand, S.; Maitre, P.; Sharma, A.; Martens, J. K.; Nicol, E.; Clavaguera, C.; Ohanessian, G. *Int. J. Mass Spectrom.* **2011**, *308*, 338-347.
- (18) Atkins, C. G.; Banu, L.; Rowsell, M.; Blagojevic, V.; Bohme, D. K.; Fridgen, T. D. *J. Phys. Chem. B* **2009**, *113*, 14457-14464.
- (19) Michaux, C.; Wouters, J.; Jacquemin, D.; Perpète, E. A. *Chem. Phys. Lett.* **2007**, *445*, 57-61.
- (20) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 2430-2438.
- (21) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 9896-9902.
- (22) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 7702-7705.
- (23) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 632-638.
- (24) Campo-Cacharrón, A.; Cabaleiro-Lago, E. M.; Rodríguez-Otero, J. *Theor Chem. Acc.* **2012**, *131*, 1290-1302.
- (25) Demireva, M.; O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, *134*, 11216-11224.
- (26) Rodziewicz, P.; Doltsinis, N. L. *ChemPhysChem* **2007**, *8*, 1959-1968.
- (27) Chang, T. M.; Prell, J. S.; Warrick, E. R.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, *134*, 15805-15813.
- (28) Bush, M. F.; Prell, J. S.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2007**, *129*, 13544-13553.
- (29) Jockusch, R. A.; Lemoff, A. S.; Williams, E. R. *J. Phys. Chem. A* **2001**, *105*, 10929-10942.
- (30) Bachrach, S. M.; Nguyen, T. T.; Demoin, D. W.; *J. Phys. Chem. A* **2009**, *113*, 6172-6181.
- (31) Kim, J. Y.; Ahn, D. S.; Park, S. W.; Lee, S. *RSC Adv.* **2014**, *4*, 16352-13361.
- (32) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. *Science* **2012**, *336*, 320-323.
- (33) Yao, Y.; Chen, D.; Zhang, S.; Li, Y.; Tu, P.; Liu, B.; Dong, M. *J. Chem. Phys. B* **2011**, *115*, 6213-6221.
- (34) Wei, Z.; Chen, D.; Zhao, H.; Li, Y.; Zhu, J.; Liu, B. *J. Chem. Phys.* **2014**, *140*, 085103 1-10.
- (35) Aikens, C. M.; Gordon, M. S. *J. Am. Chem. Soc.* **2006**, *128*, 12835-12850.
- (36) Calvo, F.; Douady, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3404-3414.
- (37) Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Chem. Phys. A* **2005**, *109*, 1903-1910.
- (38) Kamariotis, A.; Boyarkin, O.; Mercier, S.; Beck, R.; Bush, M.; Williams, E.; Rizzo, T. *J. Am. Chem. Soc.* **2006**, *128*, 905-916.
- (39) Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774-1788.
- (40) Ren, P. Y.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497-1506.

- (41) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (42) Riffet, V.; Frison, G.; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18561–18580.
- (43) Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C.; *J. Phys. Chem. B* **2012**, *116*, 483–490.
- (44) Semrouni, D.; Ohanessian, G.; Clavaguéra, C.; *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450–3462.
- (45) Jacquemin, D.; Michaux, C.; Perpète, E. A.; Frison, G. *J. Phys. Chem. B* **2011**, *115*, 3604–3613.
- (46) Addario, V.; Guo, Y.; Chu, I. V.; Ling, Y.; Ruggerio, G.; Rodriguez, C. F.; Hopkinson, A. C.; Siu, K. W. M. *Int. J. Mass Spectrom.* **2002**, *219*, 101–114.
- (47) Bouchoux, G. *Mass Spectrom. Rev.* 2014, doi: 10.1002/mas.21416.
- (48) Bouchoux, G. *Mass Spectrom. Rev.* **2012**, *31*, 391–435.
- (49) Riffet, V.; Bouchoux, G. *Phys. Chem. Chem. Phys.* 2013, *15*, 6097–6106.
- (50) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT, **2010**.
- (52) Ponder, J. W. TINKER – Software Tools for Molecular Design, version 5.0, 2009.
- (53) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- (54) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.P.; Beratan, D.N.; Yang, W. *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
- (55) Klassen, J. S.; Blades, A. T.; Kebarle, P. *J. Phys. Chem.* **1995**, *99*, 15509–15517.
- (56) Liu, D.; Wyttenbach, T.; Bowers, M.T. *Int. J. Mass Spectrom.* **2004**, *236*, 81–90.

### ***III.1.2.2. Amélioration de l'approche hiérarchique***

Notre projet s'appuyait sur l'utilisation, pour les systèmes  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  et  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$ , du protocole d'exploration de la surface d'énergie potentielle développé et testé pour le système  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$ . Cependant, comme nous le décrivons ci-après, les premiers résultats obtenus nous ont conduits à modifier notre protocole pour le rendre plus efficace.

Le protocole mis en place, s'il améliore sensiblement l'exhaustivité des approches « standards » trouvées dans la littérature, n'est pas parfait, non seulement parce qu'il demande des temps humain (construction de l'arbre de Darwin) et de calcul (optimisation de nombreuses structures en DFT) importants, mais surtout car il ne précise pas qu'elle est la fenêtre d'énergie à considérer en AMOEBA pour, au plus haut niveau de calcul considéré, obtenir l'ensemble des structures comprises dans la fenêtre d'enthalpie libre relative 0-7.5 kJ mol<sup>-1</sup>. En effet, à l'imprécision sur les énergies relatives AMOEBA vient s'ajouter le problème de l'entropie : des structures de faible énergie relative peuvent avoir une enthalpie libre relative élevée, ou inversement, comme le montrent les exemples de **C{bf}d** ( $\Delta E^\circ_{298} = 0.0$  et  $\Delta G^\circ_{298} = 14.2$  mol<sup>-1</sup>) et **Abdh** ( $\Delta E^\circ_{298} = 20.4$  et  $\Delta G^\circ_{298} = 4.5$  kJ mol<sup>-1</sup>) dans le cas de  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_3$  (voir table 3 du manuscrit ci-dessus). En conséquence, plus la fenêtre d'énergie relative AMOEBA est large, plus l'exhaustivité sera atteinte. La conséquence serait alors une augmentation importante du nombre de structures à considérer. La Figure III.1.3 illustre ce problème dans le cas de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$ , système dont la flexibilité est fortement augmentée par rapport au cas de l'histidine du fait de la chaîne latérale (CH<sub>2</sub>)<sub>4</sub>.

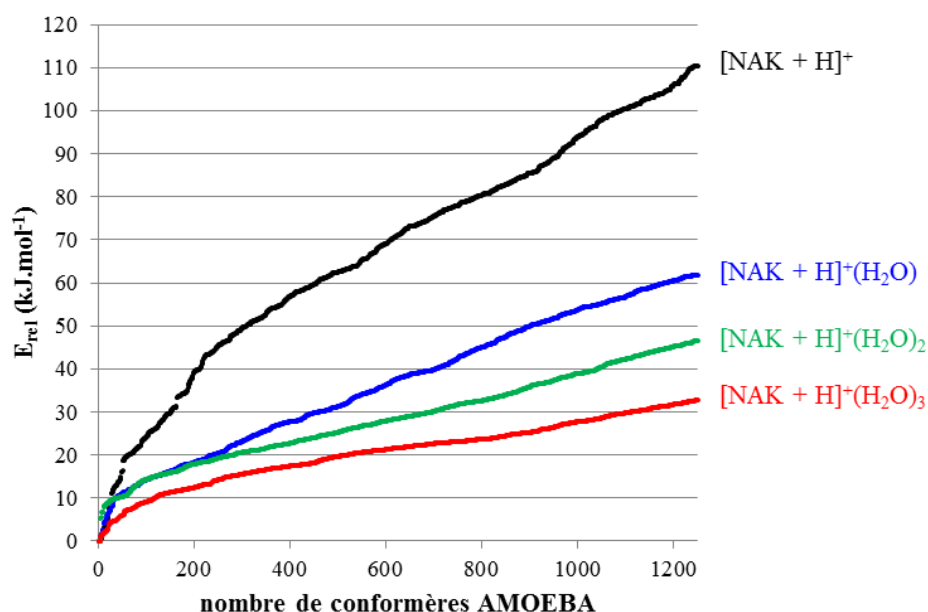


Figure III.1.3. Nombre de conformères obtenus au niveau AMOEBA, par énergie relative croissante, pour  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$ .

Plus le nombre de molécules d'eau augmente, plus la flexibilité du système augmente et plus le nombre de conformères dans une fenêtre d'énergie augmente, pour arriver par exemple à un nombre supérieur à 1000 conformères pour  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_3$  si on considère une fenêtre de 0-30  $\text{kJ mol}^{-1}$ .

Afin de mieux établir ce critère de sélection de la fenêtre d'énergie relative AMOEBA à considérer au cours de l'approche hiérarchique, différents tests ont été réalisés.

Dans un premier temps, nous avons considéré toutes les structures dans une gamme de 0-30  $\text{kJ mol}^{-1}$  pour  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n < 3$ ), et celles-ci ont été optimisées au niveau M06/6-311++G(d,p), conduisant à la détermination de nos structures « de référence » au niveau M06, et en particulier le minimum global en énergie pour chaque état d'hydratation (Tableau III.1.2). Ensuite, différents critères de sélection ont été étudiés, montrant que le nombre de conformères à considérer est moindre, mais que le minimum global en énergie n'est, dans certains cas, pas atteint, ce qui est problématique. En particulier, on constate que pour des fenêtres d'énergie de 0-10 ou 0-20  $\text{kJ mol}^{-1}$ , le minimum global de la première génération  $\mathbf{G}_0$ , i.e. pour  $[\text{NAK} + \text{H}]^+$ , n'est pas obtenu. Ceci entraîne que l'arbre de Darwin construit sur cette base serait fortement tronqué, et rien ne garantit alors qu'il sera possible de le reconstruire entièrement, donc d'avoir une recherche exhaustive des conformères de basse enthalpie libre.

Tableau III.1.2. Nombre de conformères à considérer pour les optimisations de géométrie au niveau M06/6-311++G(d,p) en fonction du critère de sélection considéré.

$E_{\text{rel}}^{\text{max-AMOEBA } a}$	$E_{\text{rel}}^{\text{max-PM6 } b}$	$[\text{NAK} + \text{H}]^+$	$[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})$	$[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_2$
10	-	<b>25<sup>c</sup></b>	<b>34<sup>c</sup></b>	35
20	-	<b>59<sup>c</sup></b>	234	277
30	-	135	450	691
10	10	44	43	41
20	10	71	236	277
20	20	94	242	277

<sup>a</sup> énergies relatives comprises entre 0 et  $E_{\text{rel}}^{\text{max-AMOEBA}}$  au niveau AMOEBA ; <sup>b</sup> énergies relatives comprises entre 0 et  $E_{\text{rel}}^{\text{max-PM6}}$  au niveau PM6-DH2//AMOEBA ; <sup>c</sup> le minimum global M06/6-311++G(d,p) en énergie n'est pas obtenu après optimisation au départ de ces seuls conformères.

L'examen attentif du minimum global en énergie de  $[\text{NAK} + \text{H}]^+$  indique que ce conformère présente une structure contrainte, avec la présence de deux liaisons hydrogène entre le groupe ammonium et les groupes C=O de l'amide et du COOH terminal (Figure III.1.4). Cette conformation particulière pourrait expliquer que AMOEBA, avec les paramètres que nous avons utilisés, donne une énergie relative élevée pour cette structure ( $\Delta E = 20.3 \text{ kJ mol}^{-1}$ ). Un examen détaillé des structures et de leurs énergies relatives semble confirmer cette constatation : seuls les conformères de structure proche du minimum global présentent un problème important d'énergie relative avec AMOEBA par rapport à M06/6-311++G(d,p).

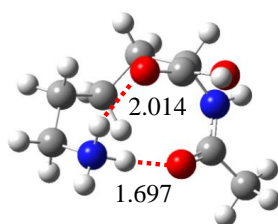


Figure III.1.4. Minimum global M06/6-311++G(d,p) en énergie pour  $[\text{NAK} + \text{H}]^+$ .

Face à cette difficulté de calcul, dans un deuxième temps, nous avons décidé de compléter notre approche par une étape supplémentaire au niveau semi-empirique. En particulier, nous avons sélectionné la méthode PM6-DH2 pour laquelle, sur la base de la paramétrisation PM6, l'inclusion récente de termes correctifs pour les liaisons H et la dispersion conduit à une bonne précision pour des temps de calculs très réduits par rapport à la DFT.[17-19] Pour toutes les structures obtenues avec AMOEBA, nous avons calculé leur énergie PM6-DH2. Trois cas se présentent alors pour une structure donnée :

- 1) Son énergie relative AMOEBA est inférieure ou égale à une valeur fixée ( $E_{\text{rel}}^{\text{max-AMOEBA}}$ ) ;
- 2) Son énergie relative AMOEBA est supérieure à une valeur fixée ( $E_{\text{rel}}^{\text{max-AMOEBA}}$ ) et son énergie relative PM6-DH2/AMOEBA est inférieure ou égale à une autre valeur fixée ( $E_{\text{rel}}^{\text{max-PM6}}$ ) ;
- 3) Son énergie relative AMOEBA est supérieure à une valeur fixée ( $E_{\text{rel}}^{\text{max-AMOEBA}}$ ) et son énergie relative PM6-DH2/AMOEBA est supérieure à une autre valeur fixée ( $E_{\text{rel}}^{\text{max-PM6}}$ ).

Notre protocole initial ne considérait que les structures appartenant au premier cas, ce qui n'était pas satisfaisant. En considérant également les structures relevant du deuxième cas (Tableau III.1.2), le minimum global est obtenu dans tous les cas pour un nombre total limité de structures. L'examen détaillé des résultats nous a de plus indiqué que des fenêtres d'énergie 0-20 kJ mol<sup>-1</sup> (i.e.  $E_{\text{rel}}^{\text{max-AMOEBA}} = E_{\text{rel}}^{\text{max-PM6}} = 20 \text{ kJ mol}^{-1}$ ) étaient nécessaires pour obtenir l'ensemble des minima situés en enthalpie libre dans la gamme 0-15 kJ mol<sup>-1</sup> au niveau M06/6-311++G(d,p). Ainsi, cette étape supplémentaire « rattrape » les minima qui seraient incorrectement décrits au niveau AMOEBA, pour un temps modeste (calculs d'énergie PM6-DH2) et sans ajouter un nombre important de nouvelles structures à ajouter à l'étape suivante (optimisation de géométrie M06/6-311++G(d,p)).

Dans un troisième temps, ce succès de PM6-DH2 nous a conduits à envisager d'autres approches utilisant les méthodes semi-empiriques. Nous avons envisagé, pour pallier d'éventuelles déficiences des géométries obtenues avec AMOEBA, d'optimiser les structures au niveau PM6 et de calculer leur énergie relative par des calculs PM6-DH2//PM6 (voir partie II.5.2 du chapitre II). Deux cas ont été considérés, soit une optimisation des géométries à partir des structures optimisées AMOEBA (pour améliorer les géométries de départ pour les calculs DFT), soit une optimisation des structures obtenues par dynamique moléculaire AMOEBA, donc avant optimisation de géométrie AMOEBA (pour « rattraper » des structures qui ne seraient pas des minima locaux sur la surface d'énergie AMOEBA). Cependant, aucune de ces approches n'a permis d'obtenir des résultats différents ou meilleurs que ce que nous avons déjà obtenu. Cela s'explique lorsque l'on compare les géométries des différentes structures obtenues (Tableau III.1.3).

Tableau III.1.3. Valeurs calculées du RMSD (Root Mean Squared Deviaion) pour les géométries AMOEBA et PM6 par rapport aux géométries M06/6-311++G(d,p), pour les 12 minima les plus stables de  $[\text{NAK} + \text{H}]^+$ .

structures	RMSD (Å)	
	AMOEBA	PM6
1	0.213	0.617
2	0.205	0.461
3	0.177	0.366
4	0.468	0.308
5	0.476	0.653
6	0.284	0.676
7	0.270	0.558
8	0.304	0.839
9	0.171	0.315
10	0.272	0.571
11	0.491	0.674
12	0.317	0.454
<b>moyenne</b>	<b>0.304</b>	<b>0.541</b>

Pour les douze structures considérées, la géométrie AMOEBA est plus proche de la géométrie DFT que la géométrie PM6, avec une erreur moyenne pratiquement divisée par 2 (0.304 Å contre 0.541 Å). En conséquence, les optimisations de géométrie PM6 n'ont pas été utilisées dans la suite de ce travail.

Suite à ce travail exploratoire, et en guise de bilan, la stratégie de modélisation que nous avons utilisée dans la suite de ce travail, pour l'exploration des surfaces d'énergie potentielle, est résumé dans la Figure III.1.5.



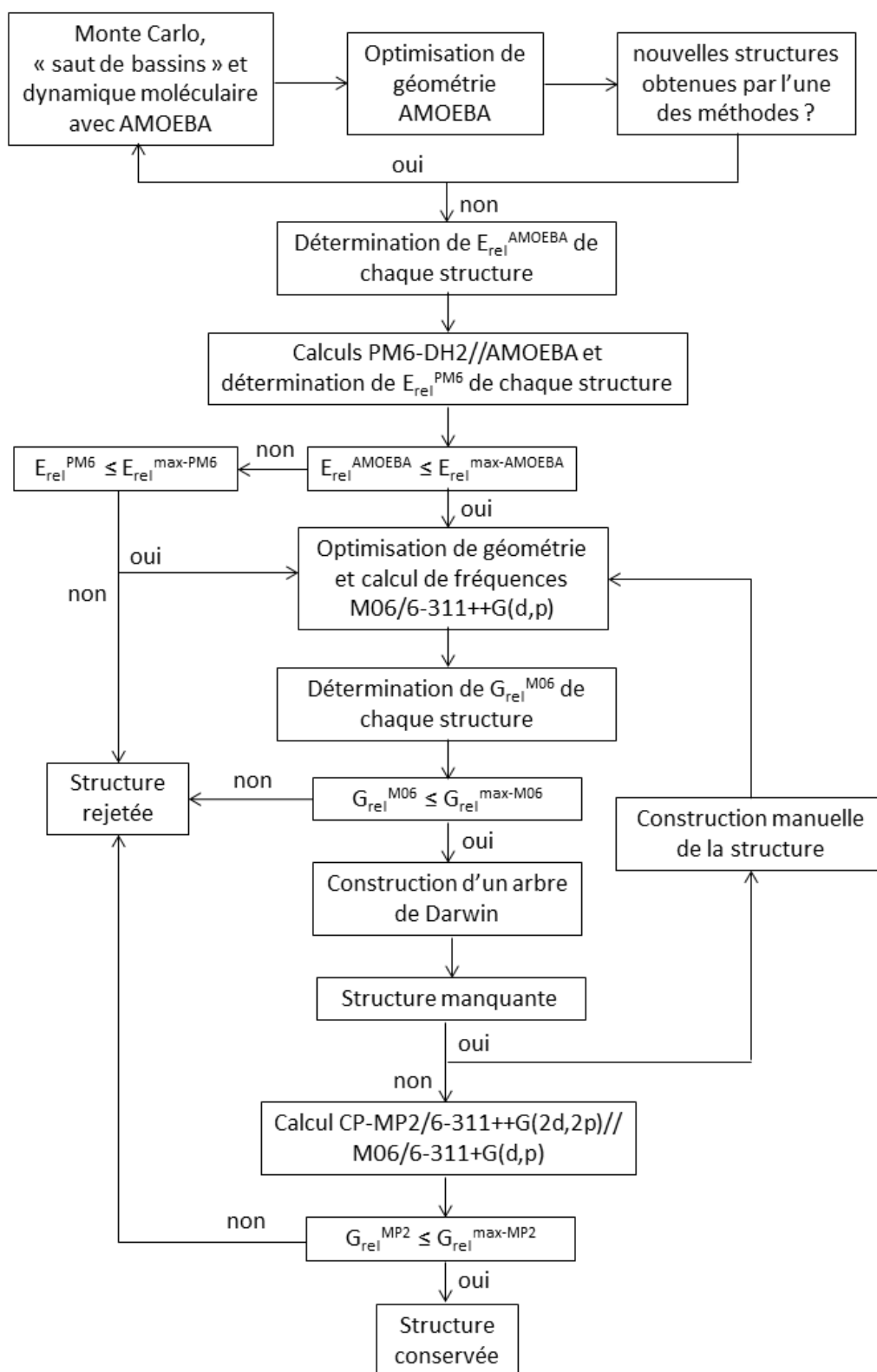


Figure III.1.5. Stratégie de modélisation mise en place pour explorer exhaustivement la surface d'énergie potentielle de systèmes flexibles. Pour l'étude des acides aminés N $\alpha$ -acétylés, les critères de sélection suivants ont été appliqués :  $E_{rel}^{max-AMOEBA} = E_{rel}^{max-PM6} = 20 \text{ kJ mol}^{-1}$  ;  $G_{rel}^{max-M06} = 15 \text{ kJ mol}^{-1}$  ;  $G_{rel}^{max-MP2} = 7.5 \text{ kJ mol}^{-1}$ .

Après l'élaboration de notre stratégie d'exploration de la surface d'énergie potentielle, nous avons complété nos résultats déjà obtenus pour  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$  (calculs CP-MP2/6-311++G(2d,2p) ) et  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  (construction de l'arbre de Darwin, ajout manuel de structures manquantes, et calculs finaux au niveau CP-MP2/6-311++G(2d,2p) ), et appliqué intégralement cette stratégie à  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$ . Sur l'ensemble de ces calculs, nous pouvons estimer que, parmi les structures situées dans la gamme d'enthalpie libre 0-15 kJ mol<sup>-1</sup> en M06/6-311++G(d,p), environ 60% ont été identifiées lors de l'approche hiérarchique alors que les autres structures sont obtenues par construction de l'arbre de Darwin. La description détaillée de la recherche conformationnelle pour les systèmes  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  et  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$  n'est pas présentée dans ce manuscrit car elle n'apporte pas de nouvelles conclusions sur notre approche. Le tableau III.1.4 résume l'ensemble des structures non-hydratées situées dans la gamme d'enthalpie libre 0-15 kJ mol<sup>-1</sup> au niveau M06/6-311++G(d,p), ainsi que leurs enthalpies libres relatives CP-MP2/6-311++G(2d,2p)/M06/6-311++G(d,p), qui sont utilisées pour les classer et les numéroter. Leur nom comprend également la lettre correspondant à l'acide aminé et l'indice « 0 » pour indiquer que ce sont des systèmes sans molécule d'eau (n = 0). Une attention particulière est portée sur cette première génération car elle est cruciale pour la construction de l'arbre de Darwin. L'accord entre les deux niveaux de calcul est raisonnable, même si des écarts de plusieurs kJ mol<sup>-1</sup> sont observés, ce qui justifie la fenêtre 0-15 kJ mol<sup>-1</sup> en DFT pour garantir l'exhaustivité dans la fenêtre 0-7.5 kJ mol<sup>-1</sup> au plus haut niveau de calcul utilisé. Les structures précédemment obtenues pour  $[\text{NAH} + \text{H}]^+$  avec M06 (voir Figure 2 de l'article ci-dessus) sont identiques, mais dans un ordre légèrement différent au niveau de calcul MP2 (Figure III.1.6).

Tableau III.1.4. Ensemble des conformations de  $[\text{NAK} + \text{H}]^+$ ,  $[\text{NAR} + \text{H}]^+$  et  $[\text{NAH} + \text{H}]^+$  situées dans la gamme d'enthalpie libre de 0-15 kJ/mol au niveau M06/6-311++G(d,p).

$[\text{NAK} + \text{H}]^+$	Type	$\Delta G_{298}^\circ$		$[\text{NAR} + \text{H}]^+$	Type	$\Delta G_{298}^\circ$		$[\text{NAH} + \text{H}]^+$	$\Delta G_{298}^\circ$	
		M06 <sup>a</sup>	MP2//M06 <sup>b</sup>			M06 <sup>a</sup>	MP2//M06 <sup>b</sup>		M06 <sup>a</sup>	MP2//M06 <sup>b</sup>
<b>K<sup>0</sup>_1</b>	I	0.1	0.0	<b>R<sup>0</sup>_1</b>	I	0.0	0.0	<b>H<sup>0</sup>_1</b>	0.0	0.0
<b>K<sup>0</sup>_2</b>	I	2.0	0.3	<b>R<sup>0</sup>_2</b>	I	1.1	2.9	<b>H<sup>0</sup>_2</b>	6.7	4.3
<b>K<sup>0</sup>_3</b>	I	3.0	2.5	<b>R<sup>0</sup>_3</b>	I	4.0	8.5	<b>H<sup>0</sup>_3</b>	7.1	6.0
<b>K<sup>0</sup>_4</b>	II	3.2	3.3	<b>R<sup>0</sup>_4</b>	II	7.2	9.2	<b>H<sup>0</sup>_4</b>	5.7	6.9
<b>K<sup>0</sup>_5</b>	III	0.0	3.5	<b>R<sup>0</sup>_5</b>	I	8.0	9.7	<b>H<sup>0</sup>_5</b>	12.9	10.7
<b>K<sup>0</sup>_6</b>	I	2.3	3.6	<b>R<sup>0</sup>_6</b>	III	10.3	11.8			
<b>K<sup>0</sup>_7</b>	II	9.4	8.8	<b>R<sup>0</sup>_7</b>	I	11.5	12.6			
<b>K<sup>0</sup>_8</b>	I	9.9	9.2	<b>R<sup>0</sup>_8</b>	III	12.5	15.4			
<b>K<sup>0</sup>_9</b>	II	8.7	10.0	<b>R<sup>0</sup>_9</b>	I	12.0	16.0			
<b>K<sup>0</sup>_10</b>	II	11.3	10.1	<b>R<sup>0</sup>_10</b>	I	11.3	16.2			
<b>K<sup>0</sup>_11</b>	III	10.5	10.2	<b>R<sup>0</sup>_11</b>	III	11.7	16.4			
<b>K<sup>0</sup>_12</b>	I	12.4	10.6							
<b>K<sup>0</sup>_13</b>	I	9.1	11.4							
<b>K<sup>0</sup>_14</b>	I	8.8	12.3							
<b>K<sup>0</sup>_15</b>	I	12.8	13.0							
<b>K<sup>0</sup>_16</b>	I	11.5	13.1							
<b>K<sup>0</sup>_17</b>	III	14.6	17.4							

<sup>a</sup> Enthalpies libres relatives calculées au niveau M06/6-311++G(d,p). <sup>b</sup> Enthalpies libres relatives calculées au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

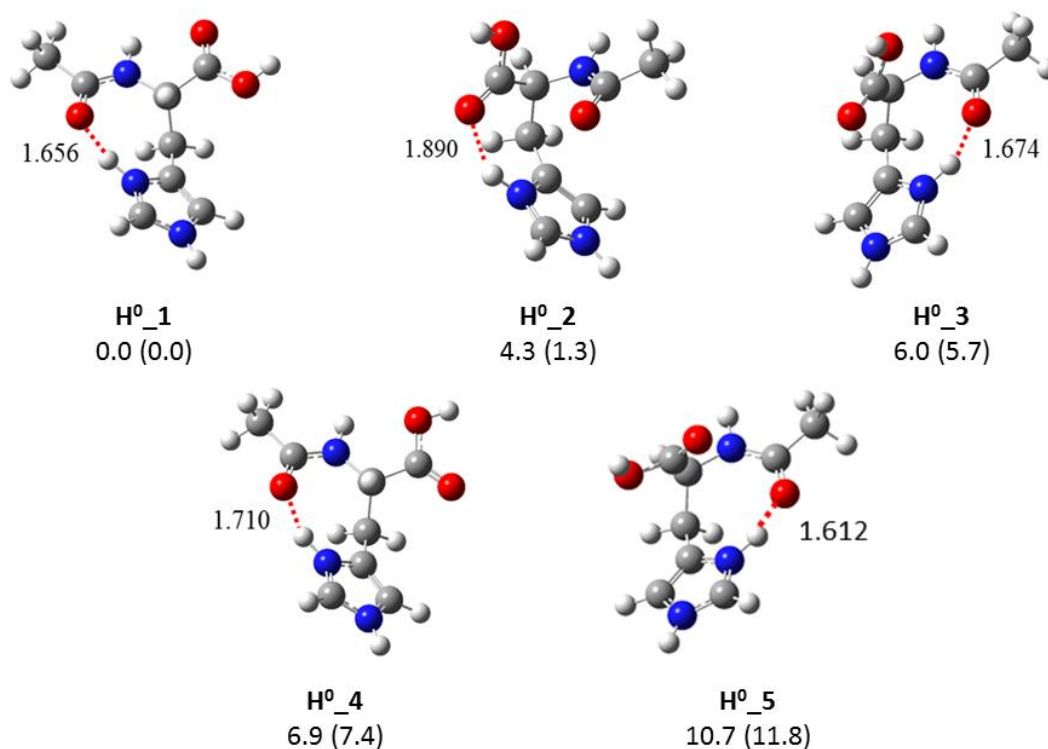


Figure III.1.6. Structures optimisées au niveau M06/6-311++G(d,p) des 5 conformères les plus stables de  $[\text{NAH} + \text{H}]^+$ . Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

Comme noté précédemment, ces 5 structures présentent des conformations où l'orientation relative des groupes imidazole, COOH et amide est différente. Une situation différente est observée pour  $[\text{NAK} + \text{H}]^+$  et  $[\text{NAR} + \text{H}]^+$  (Figures III.1.7 et III.1.8). En effet, du fait de la flexibilité de leur chaîne latérale, des conformations très proches sont parfois obtenues, si l'on excepte la conformation de la chaîne alkyle. Dans chaque cas, nous avons donc classé ces structures en trois types de conformations. Au sein de chaque type, seule la conformation de la chaîne alkyle change.

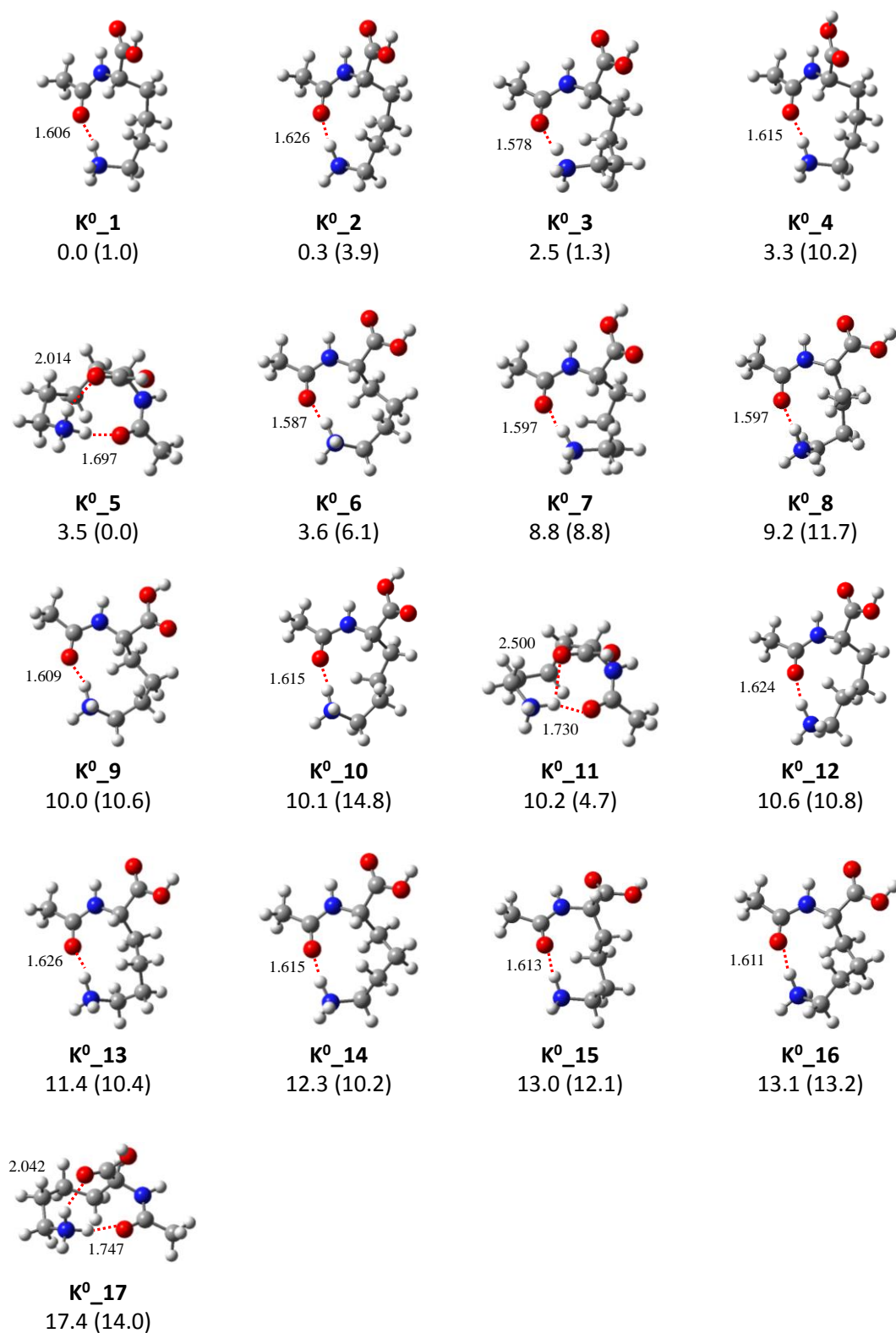


Figure III.1.7. Structures optimisées au niveau M06/6-311++G(d,p) des 17 conformères les plus stables de  $[\text{NAK} + \text{H}]^+$ . Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

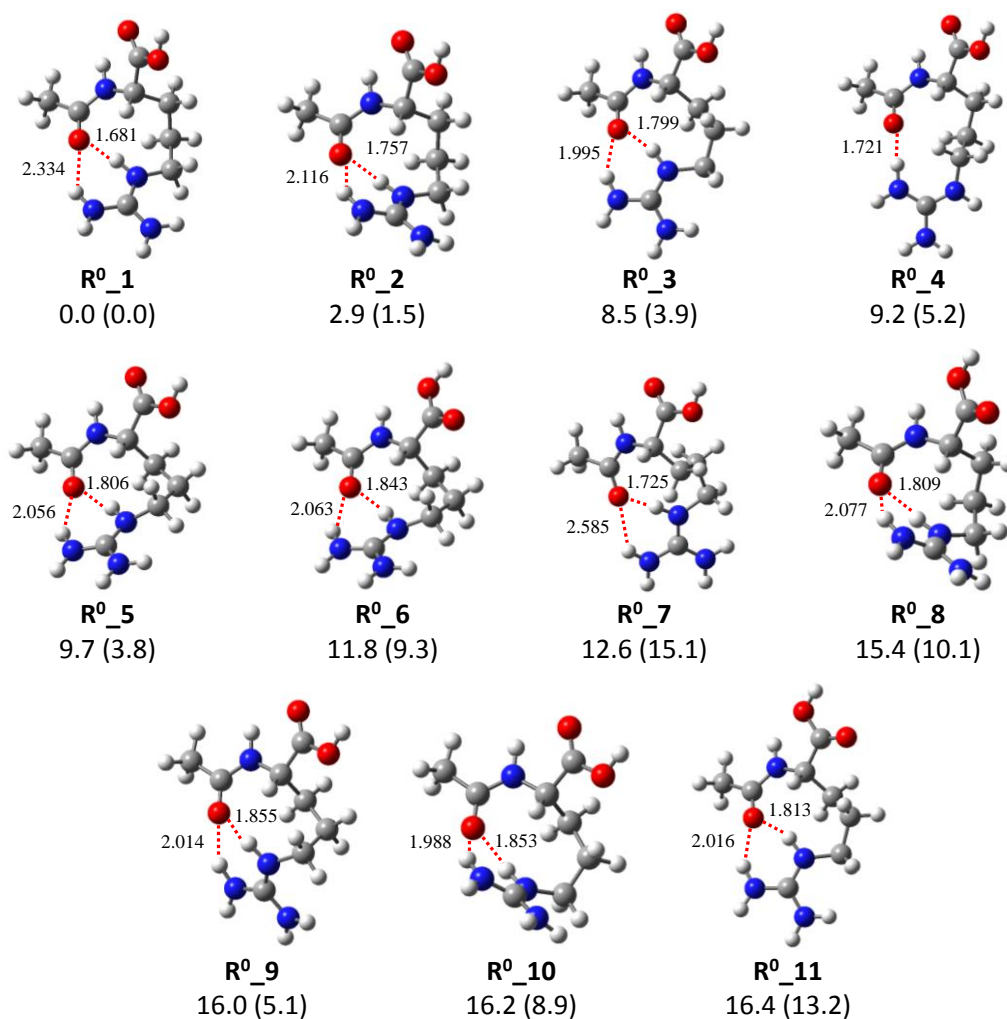


Figure III.1.8. Structures optimisées au niveau M06/6-311++G(d,p) des 11 conformères les plus stables de  $[\text{NAR} + \text{H}]^+$ . Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

Toutes ces structures ont une conformation repliée du fait de la présence de liaisons hydrogène intramoléculaires entre les groupes chargés et les C=O des groupes amide et/ou COOH. Suivant la conformation, l'ammonium et le guanidinium peuvent faire une ( $[\text{NAK} + \text{H}]^+$  types I et II ;  $[\text{NAR} + \text{H}]^+$  type II) ou deux liaisons hydrogène ( $[\text{NAK} + \text{H}]^+$  type III ;  $[\text{NAK} + \text{H}]^+$  types I et III). On constate que la présence de deux interactions conduit généralement à des liaisons hydrogène moins fortes (longueur de liaison plus longue) du fait de contraintes géométriques, donc entropiquement défavorables. En conséquence, même si ces deux interactions non-covalentes peuvent être favorables d'un point de vue enthalpique (par exemple **K<sup>0</sup>\_5** est le conformère de plus basse enthalpie), cela n'est plus vrai du point de vue de l'enthalpie libre.

Seules les structures situées dans la gamme d'enthalpie libre 0-7.5 kJ mol<sup>-1</sup> en CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p) ont été prises en compte dans la partie III.1.3.

### III.1.3. Influences réciproques des interactions non-covalentes

#### Structures

Le tableau III.1.5 récapitule l'ensemble des conformères de basse énergie obtenus pour les systèmes  $[\text{NAK}+\text{H}]^+(\text{H}_2\text{O})_n$ ,  $[\text{NAR}+\text{H}]^+(\text{H}_2\text{O})_n$  et  $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$  ( $n=0, 1, 2$  and  $3$ ), représentés dans les Figures III.1.9 ( $[\text{NAK}+\text{H}]^+(\text{H}_2\text{O})_n$ ), III.1.10 ( $[\text{NAR}+\text{H}]^+(\text{H}_2\text{O})_n$ ) et III.1.11 ( $[\text{NAH}+\text{H}]^+(\text{H}_2\text{O})_n$ ) (pour la génération  $\mathbf{G}_0$ , voir les Figures III.1.6, III.1.7 et III.1.8).

Tableau III.1.5. Enthalpies ( $\Delta H^\circ_{298}$ ) et enthalpies libres relatives ( $\Delta G^\circ_{298}$ ), en kJ mol<sup>-1</sup>, et fractions molaires ( $x_i$ ) calculées au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p) pour les structures, situées dans la fenêtre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre, de [NAK+H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub>, [NAR+H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub> et [NAH+H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub> (n=0, 1, 2 and 3).

[NAK+H] <sup>+</sup> (H <sub>2</sub> O) <sub>n</sub>				[NAR+H] <sup>+</sup> (H <sub>2</sub> O) <sub>n</sub>				[NAH+H] <sup>+</sup> (H <sub>2</sub> O) <sub>n</sub>			
Structures	$\Delta H^\circ_{298}$	$\Delta G^\circ_{298}$	$x_i$	Structures	$\Delta H^\circ_{298}$	$\Delta G^\circ_{298}$	$x_i$	Structures	$\Delta H^\circ_{298}$	$\Delta G^\circ_{298}$	$x_i$
<b>n = 0. Generation G<sub>0</sub></b>											
K <sup>0</sup> _1	1.0	0.0	0.34	R <sup>0</sup> _1	0.0	0.0	0.76	H <sup>0</sup> _1	0.0	0.0	0.75
K <sup>0</sup> _2	3.9	0.3	0.30	R <sup>0</sup> _2	1.5	2.9	0.24	H <sup>0</sup> _2	1.3	4.3	0.14
K <sup>0</sup> _3	1.3	2.5	0.12					H <sup>0</sup> _3	5.7	6.0	0.07
K <sup>0</sup> _4	10.2	3.3	0.09					H <sup>0</sup> _4	7.4	6.9	0.05
K <sup>0</sup> _5	0.0	3.5	0.08								
K <sup>0</sup> _6	6.1	3.6	0.06								
<b>n = 1. Generation G<sub>1</sub></b>											
K <sup>1</sup> _1	0.4	0.0	0.35	R <sup>1</sup> _1	0.0	0.0	0.42	H <sup>1</sup> _1	0.0	0.0	0.76
K <sup>1</sup> _2	0.0	0.5	0.28	R <sup>1</sup> _2	7.8	2.5	0.15	H <sup>1</sup> _2	1.5	5.1	0.10
K <sup>1</sup> _3	2.0	0.9	0.24	R <sup>1</sup> _3	2.5	2.5	0.15	H <sup>1</sup> _3	6.7	5.3	0.09
K <sup>1</sup> _4	2.4	4.3	0.06	R <sup>1</sup> _4	2.9	4.3	0.08	H <sup>1</sup> _4	6.8	6.6	0.05
K <sup>1</sup> _5	8.9	5.3	0.04	R <sup>1</sup> _5	7.7	4.5	0.07				
K <sup>1</sup> _6	2.7	6.6	0.03	R <sup>1</sup> _6	2.6	5.3	0.05				
				R <sup>1</sup> _7	1.8	6.1	0.04				
				R <sup>1</sup> _8	5.0	7.0	0.02				
				R <sup>1</sup> _9	9.2	7.2	0.02				
<b>n = 2. Generation G<sub>2</sub></b>											
K <sup>2</sup> _1	6.9	0.0	0.32	R <sup>2</sup> _1	0.3	0.0	0.52	H <sup>2</sup> _1	0.0	0.0	0.33
K <sup>2</sup> _2	4.9	0.3	0.28	R <sup>2</sup> _2	4.0	2.4	0.20	H <sup>2</sup> _2	7.8	0.8	0.24
K <sup>2</sup> _3	4.6	3.6	0.08	R <sup>2</sup> _3	4.5	3.1	0.15	H <sup>2</sup> _3	6.2	1.4	0.19
K <sup>2</sup> _4	3.3	4.9	0.05	R <sup>2</sup> _4	0.0	5.2	0.06	H <sup>2</sup> _4	6.3	3.3	0.09
K <sup>2</sup> _5	5.2	5.8	0.03	R <sup>2</sup> _5	6.9	6.7	0.04	H <sup>2</sup> _5	7.5	4.8	0.05
K <sup>2</sup> _6	5.4	5.8	0.03	R <sup>2</sup> _6	6.7	7.2	0.03	H <sup>2</sup> _6	7.6	5.7	0.03
K <sup>2</sup> _7	9.1	5.9	0.03					H <sup>2</sup> _7	16.5	6.7	0.02
K <sup>2</sup> _8	5.6	6.0	0.03					H <sup>2</sup> _8	13.1	6.8	0.02
K <sup>2</sup> _9	1.3	6.3	0.03					H <sup>2</sup> _9	5.9	6.9	0.02
K <sup>2</sup> _10	8.7	6.4	0.02								
K <sup>2</sup> _11	14.8	6.4	0.02								
K <sup>2</sup> _12	14.2	6.5	0.02								
K <sup>2</sup> _13	9.4	6.7	0.02								
K <sup>2</sup> _14	9.0	7.0	0.02								
K <sup>2</sup> _15	6.8	7.4	0.02								
<b>n = 3. Generation G<sub>3</sub></b>											
K <sup>3</sup> _1	8.0	0.0	0.44	R <sup>3</sup> _1	3.5	0.0	0.32	H <sup>3</sup> _1	2.3	0.0	0.41
K <sup>3</sup> _2	13.2	2.1	0.19	R <sup>3</sup> _2	3.3	0.2	0.29	H <sup>3</sup> _2	2.2	1.7	0.21
K <sup>3</sup> _3	2.6	2.5	0.16	R <sup>3</sup> _3	3.5	4.2	0.06	H <sup>3</sup> _3	13.1	2.6	0.14
K <sup>3</sup> _4	12.1	3.7	0.10	R <sup>3</sup> _4	2.1	5.2	0.04	H <sup>3</sup> _4	9.6	4.5	0.07
K <sup>3</sup> _5	14.1	4.9	0.07	R <sup>3</sup> _5	3.9	5.2	0.04	H <sup>3</sup> _5	14.3	4.9	0.06
K <sup>3</sup> _6	4.2	5.9	0.04	R <sup>3</sup> _6	0.1	5.4	0.04	H <sup>3</sup> _6	0.0	5.0	0.06
				R <sup>3</sup> _7	5.5	5.5	0.03	H <sup>3</sup> _7	9.6	5.5	0.05
				R <sup>3</sup> _8	6.9	5.8	0.03				
				R <sup>3</sup> _9	10.4	5.8	0.03				
				R <sup>3</sup> _10	6.1	5.9	0.03				
				R <sup>3</sup> _11	12.4	6.5	0.02				
				R <sup>3</sup> _12	0.0	6.6	0.02				
				R <sup>3</sup> _13	12.9	6.9	0.02				
				R <sup>3</sup> _14	10.4	7.5	0.02				
				R <sup>3</sup> _15	12.5	7.5	0.02				



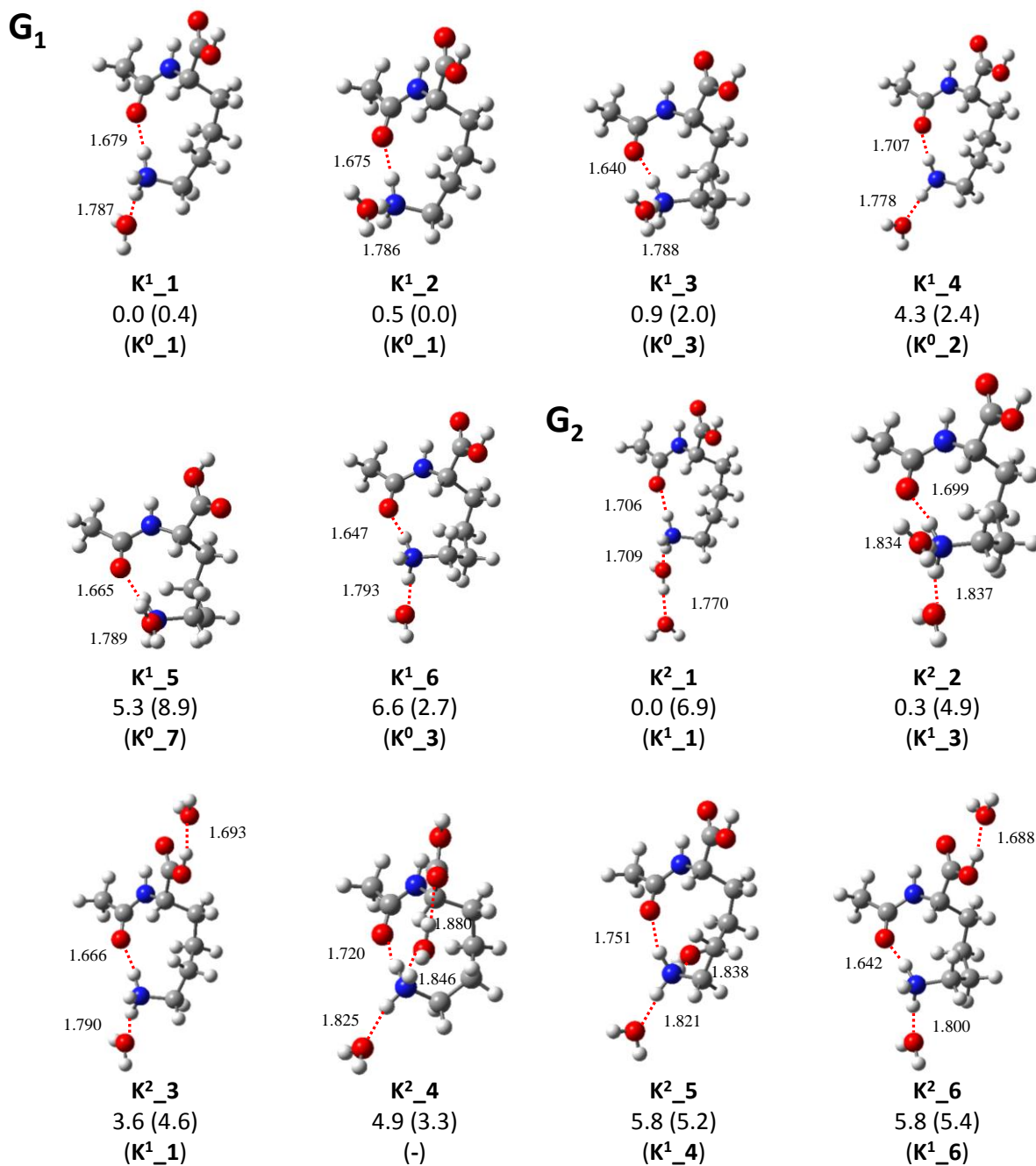


Figure III.1.9. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

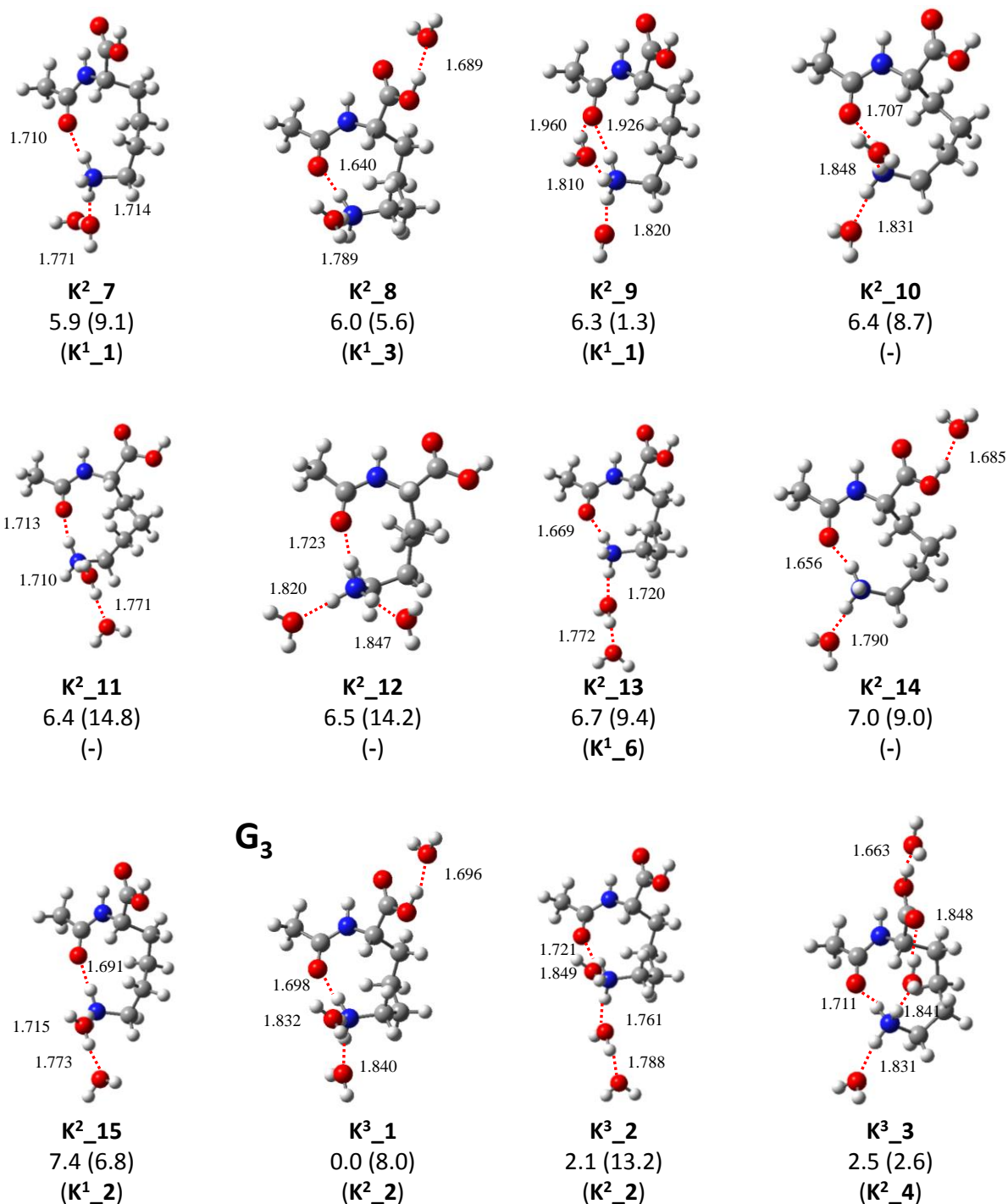


Figure III.1.9 – suite. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

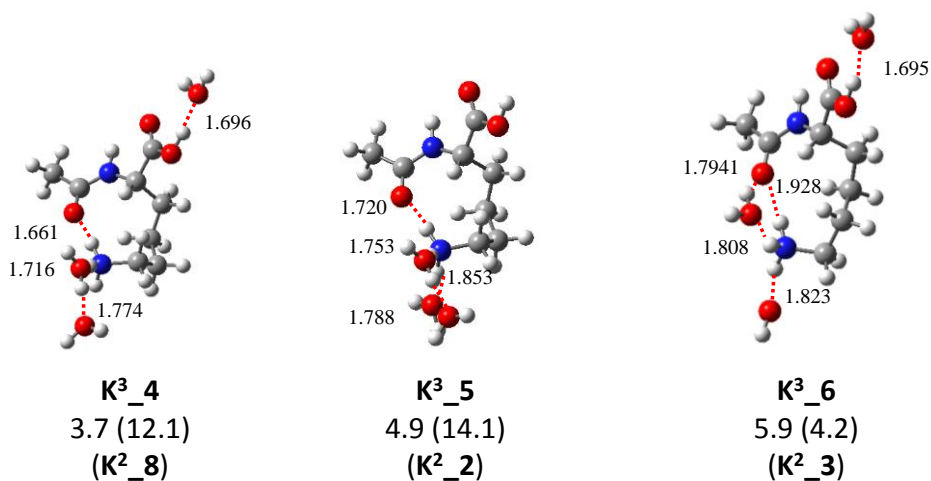


Figure III.1.9 – suite. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

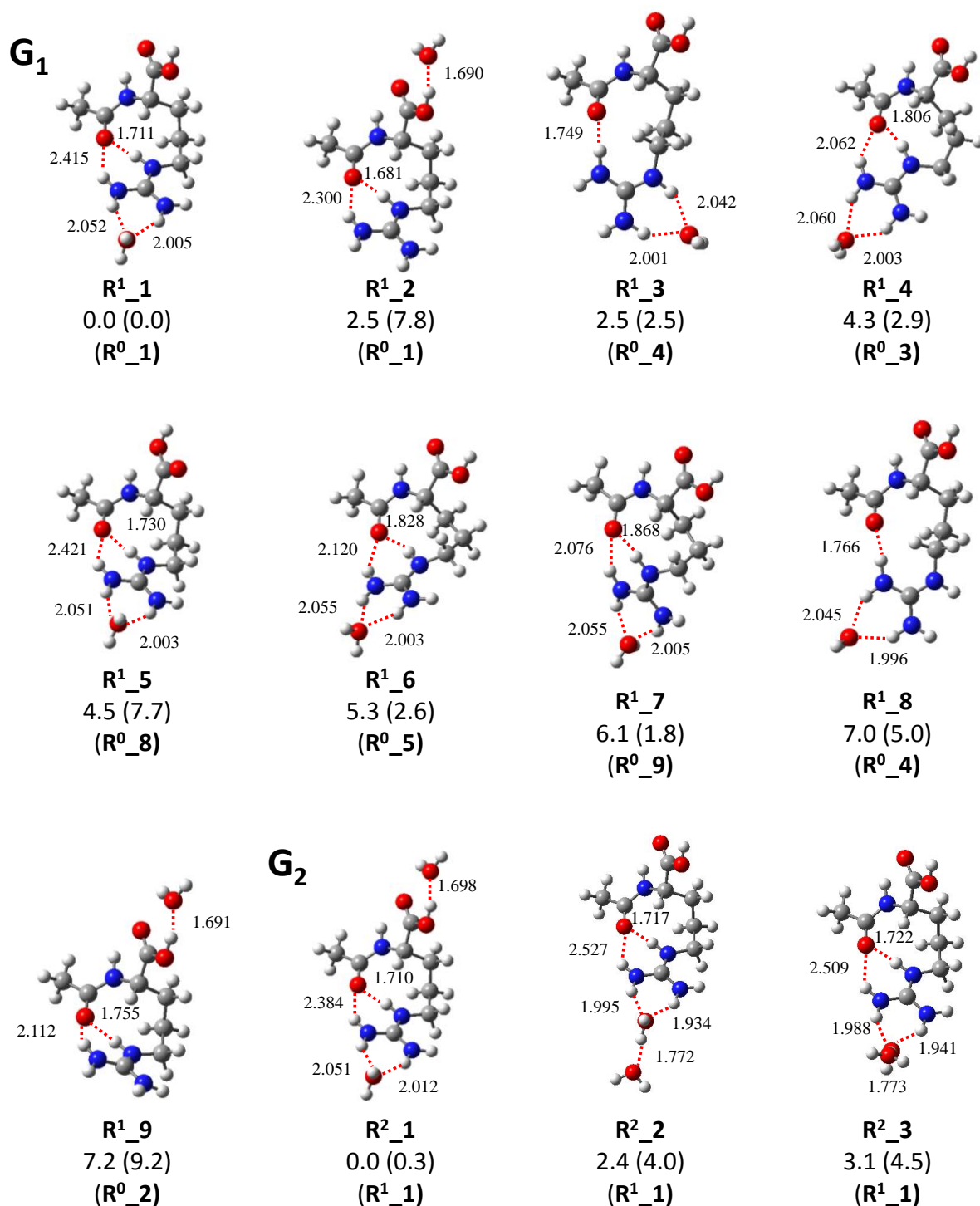


Figure III.1.10. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

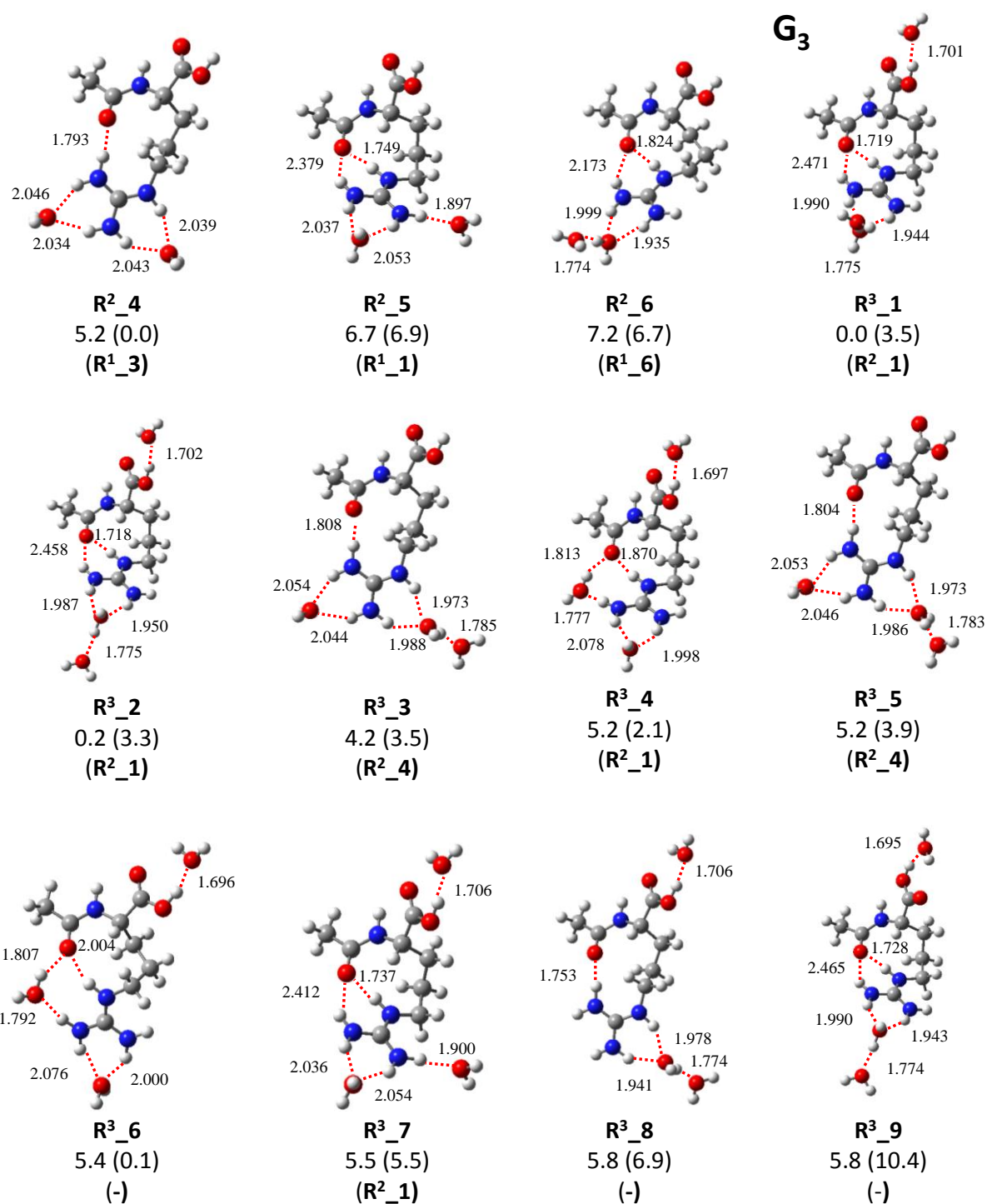


Figure III.1.10 – suite. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

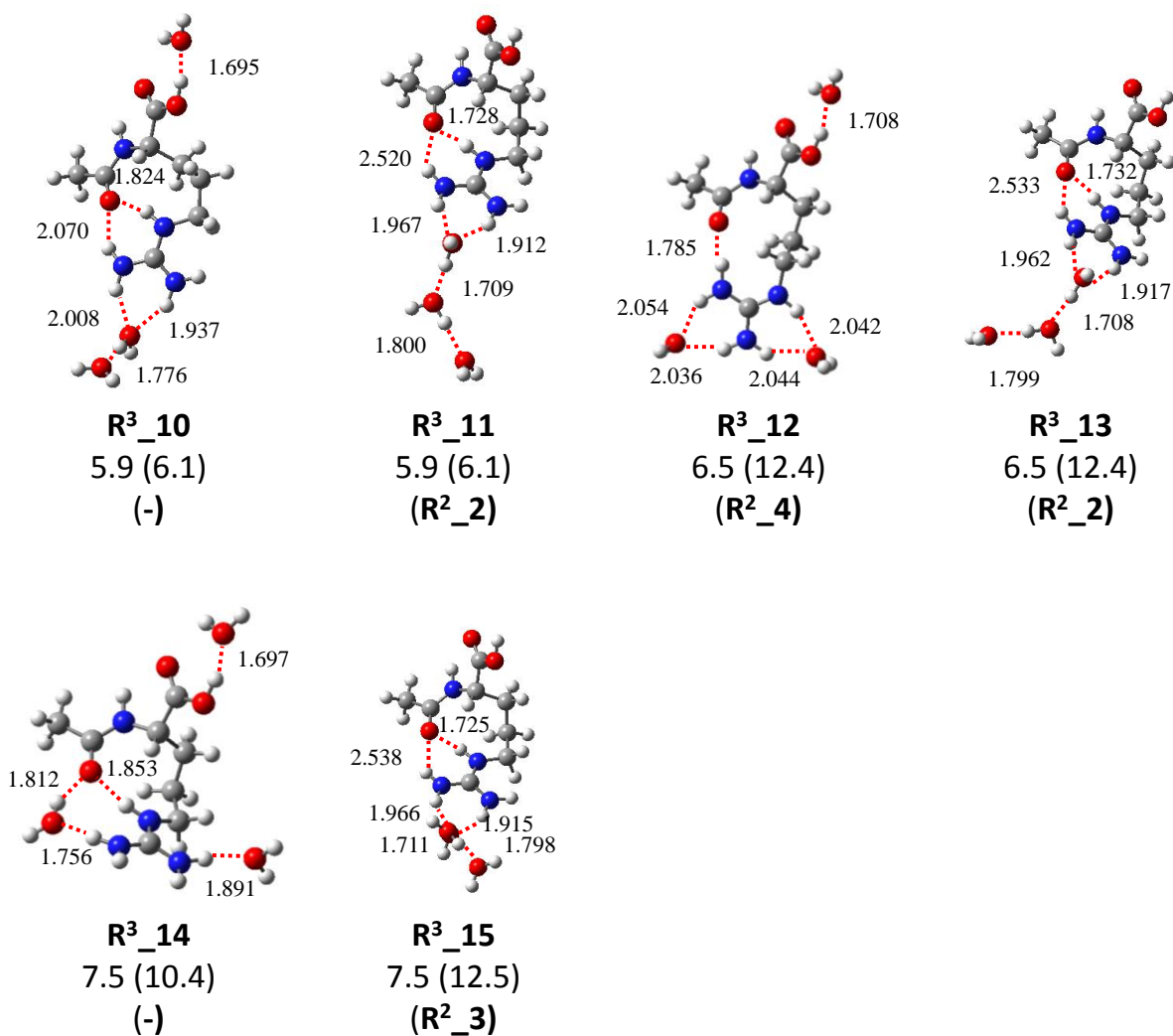


Figure III.1.10 – suite. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.



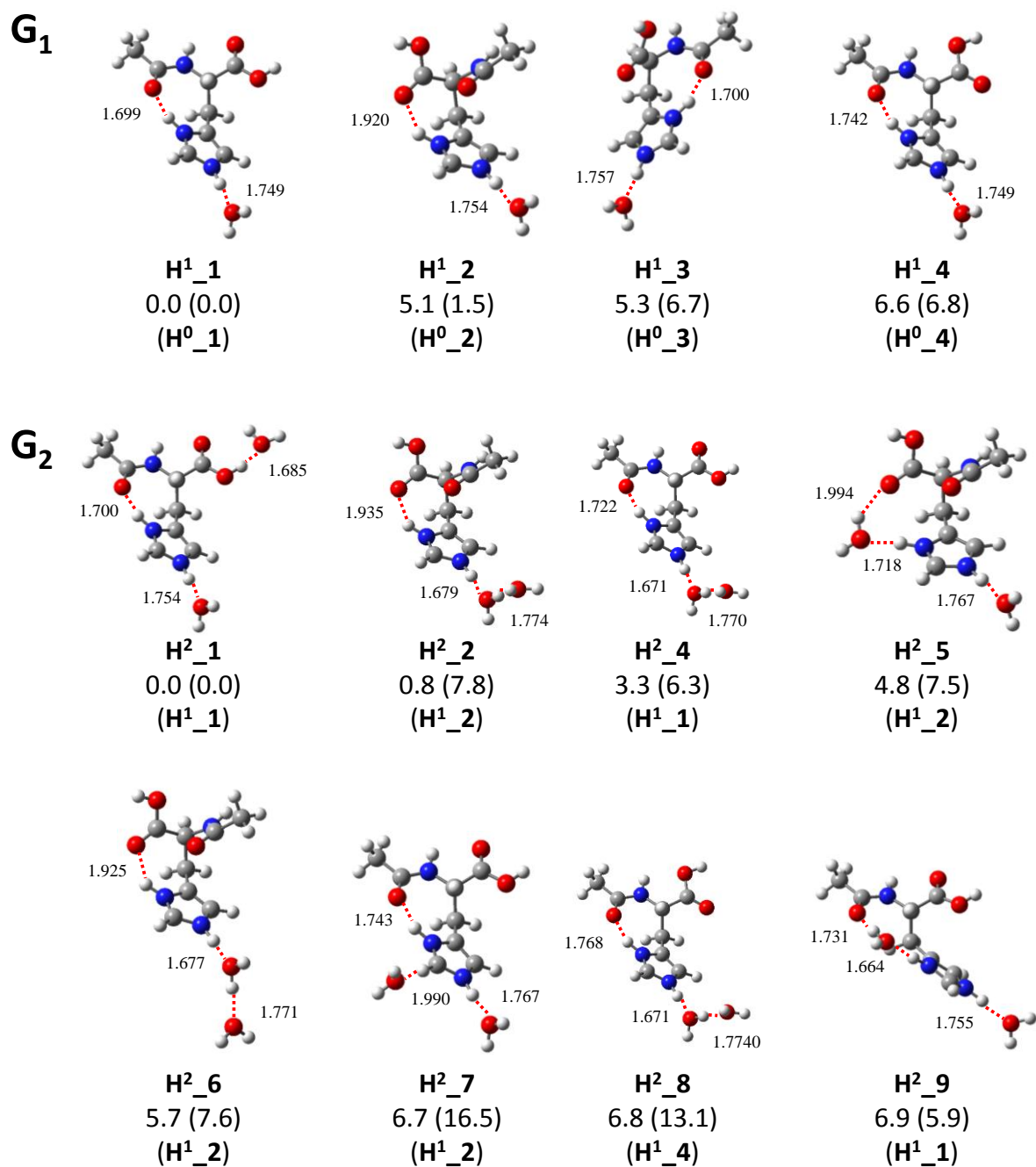


Figure III.1.11. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

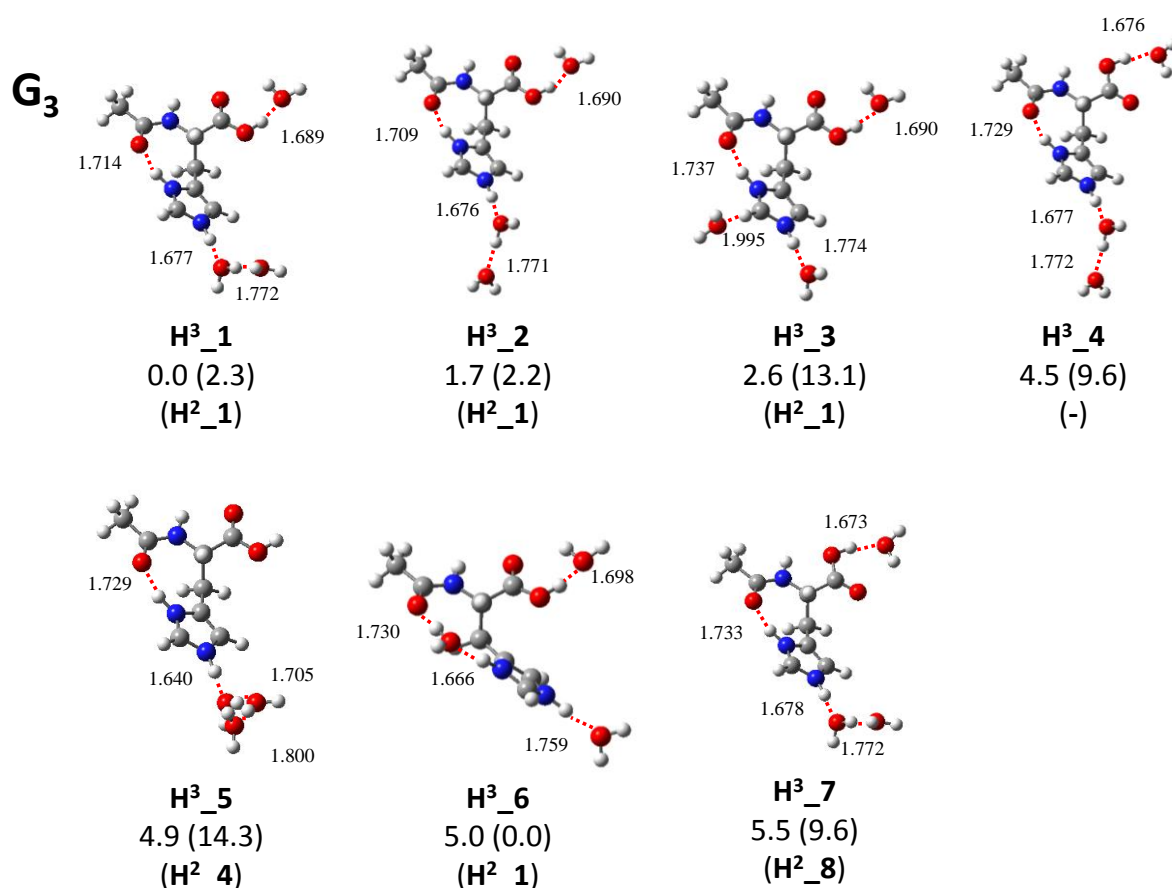


Figure III.1.11 – suite. Structures optimisées au niveau M06/6-311++G(d,p) des conformères les plus stables de  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$  ( $n = 1, 2, 3$ ). Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p). Le parent (i.e. la structure équivalente de la génération précédente) de chaque structure est indiqué entre parenthèse. La notation (-) signifie que le parent n'est pas situé entre 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre.

Contrairement à ce qui était observé pour la première génération, le nombre de conformères identifiés dans la gamme d'enthalpie libre 0-7.5 kJ mol<sup>-1</sup> ne dépend pas de la flexibilité du système. Cependant, en général, la flexibilité diminue l'écart d'énergie entre les conformères, conduisant à réduire progressivement la proportion du minimum global lors de l'ajout de molécules d'eau.

Comme nous l'avons vu précédemment dans le cas de l'histidine, le site de plus forte hydratation est une liaison NH du groupe chargé dans tous les cas. Ainsi, 100% des structures de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_1$  et  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_1$  sont dans ce cas. Dans le cas de  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_1$ , on observe néanmoins que 17% de la population d'équilibre (**R<sup>1</sup>\_2** et **R<sup>1</sup>\_9**) présente une liaison hydrogène entre la molécule d'eau et le groupe COOH. De plus, dans les autres cas, la molécule d'eau n'est pas liée à un groupe NH, mais en pont entre deux groupes NH. Dans le cas de  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_1$ , la molécule d'eau ne se place pas de façon



équiprobable sur l'un de ces deux NH du groupe chargé puisqu'un rapport 38 : 62 dans la population d'équilibre est observée (voir Tableau III.1.5). Pour expliquer ce phénomène, nous avons comparé les charges de Mulliken des deux atomes d'hydrogène disponibles (NH du groupe chargé) pour les structures parentes  $[\text{NAK} + \text{H}]^+$  pertinentes :  $\mathbf{K}^0\_1$ ,  $\mathbf{K}^0\_2$ ,  $\mathbf{K}^0\_3$  et  $\mathbf{K}^0\_7$ . La Figure III.1.12 nous permet de constater que ces deux atomes d'hydrogène ne sont effectivement pas équivalents et que la molécule d'eau interagit préférentiellement avec l'atome d'hydrogène ayant la plus grande charge positive.

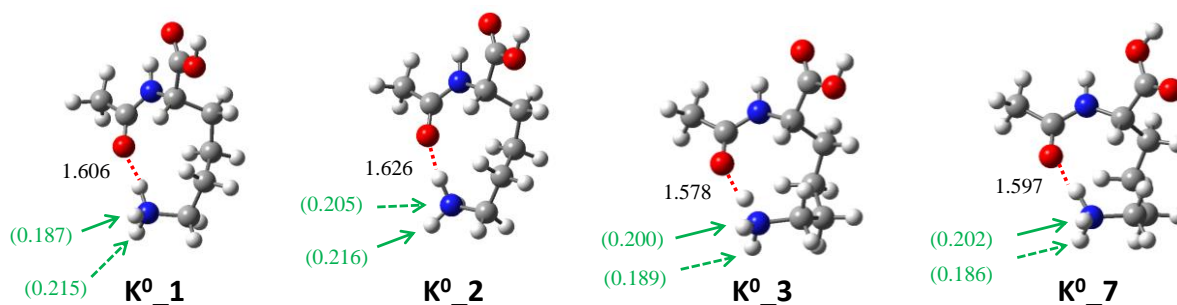


Figure III.1.12. Charges de Mulliken des atomes d'hydrogène disponible du groupe chargé de  $\mathbf{K}^0\_1$ ,  $\mathbf{K}^0\_2$ ,  $\mathbf{K}^0\_3$  et  $\mathbf{K}^0\_7$  calculées au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

Le site de fixation de la seconde et de la troisième molécule d'eau présente une plus grande diversité en fonction de la nature de l'acide aminé considéré. Dans le cas de  $[\text{NAH} + \text{H}]^+(\text{H}_2\text{O})_n$ , comme vu précédemment, l'absence d'une autre liaison NH chargée devrait conduire soit à la fixation sur le groupe COOH ou en seconde sphère d'hydratation, soit à l'insertion dans la liaison hydrogène intramoléculaire. De façon plus surprenante, dans le cas de  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})_n$ , la seconde molécule d'eau ne s'attache que minoritairement au site chargé ( $\mathbf{R}^2\_4$  et  $\mathbf{R}^2\_5$ ). On retrouve majoritairement, comme pour l'histidine, des fixations sur le groupe COOH ou en seconde sphère d'hydratation. Enfin,  $[\text{NAK} + \text{H}]^+(\text{H}_2\text{O})_n$  présente tous les cas possibles, sans préférence très marquée.

La littérature[15,20-25] indique que, si l'on ne considère que le conformère le plus stable d'une génération, celui-ci est le fils du conformère le plus stable de la génération précédente. Cette situation est retrouvée dans tous les cas, sauf un :  $\mathbf{K}^3\_1$  est le fils de  $\mathbf{K}^2\_2$ . Ceci illustre un des aspects discuté précédemment : la stratégie de l'arbre de Darwin, appliquée, seule, à des systèmes de plus en plus flexibles, pourrait ne pas permettre d'identifier des conformères issus de parents qui ne proviennent pas de la branche de plus basse énergie.

## Effets des interactions non-covalentes et données thermochimiques d'hydratation.

La micro-solvatation de ces systèmes induit différentes modifications structurales. Premièrement, si l'on considère un parent d'une conformation donnée, sa liaison hydrogène intramoléculaire n'est allongée que par l'addition d'une molécule d'eau sur le site chargé, alors que l'addition sur le groupe COOH ne la perturbe pas, comme illustré dans le Tableau III.1.6.

Tableau III.1.6. Evolution de la longueur de la liaison hydrogène intramoléculaire.

Site d'hydratation	Distance NH...OCNH (Å)		
	[NAH+H] <sup>+</sup> Parent <b>H<sup>0</sup>_1</b>	[NAK+H] <sup>+</sup> Parent <b>K<sup>0</sup>_3</b>	[NAR+H] <sup>+</sup> Parent <b>R<sup>0</sup>_1</b>
<b>G<sub>0</sub></b> -	1.656	1.578	1.681
<b>G<sub>1</sub></b> NH du groupe chargé	1.699	1.640	1.711
<b>G<sub>2</sub></b> NH du groupe chargé	/	1.699	1.749
Groupe COOH	1.700	/	1.710
<b>G<sub>3</sub></b> Groupe COOH	/	1.698	/

Ainsi, l'hydratation affaiblit l'interaction non covalente intramoléculaire, et l'insertion d'une molécule d'eau dans cette liaison hydrogène devient possible, comme observé pour **H<sup>2</sup>\_5**, **H<sup>2</sup>\_9**, **H<sup>3</sup>\_6**, **R<sup>3</sup>\_4**, **R<sup>3</sup>\_6** et **R<sup>3</sup>\_14**. Dans le cas de la lysine, nous n'avons pas observé l'insertion d'une molécule d'eau dans la liaison hydrogène intramoléculaire pour les structures de basse énergie. A l'inverse, la formation, par une molécule d'eau, de deux liaisons hydrogène avec deux sites distants est observée (**K<sup>2</sup>\_4**, **K<sup>3</sup>\_3**). Ces insertions et « pontage » par une molécule d'eau conduisent à des modifications structurales parfois significatives du squelette du parent correspondant. Néanmoins, dans chaque cas, un parent a été défini, soit du fait de la plus grande ressemblance, soit éventuellement en réalisant une optimisation de géométrie après effacement de la molécule d'eau ajoutée.

L'ensemble de nos données permet de calculer les énergies, enthalpies, entropies et enthalpies libres d'hydratation pour les différents systèmes étudiés (Tableau III.1.7).

Tableau III.1.7. Données thermochimiques d'hydratation de  $[\text{NAH}+\text{H}]^+$ ,  $[\text{NAR}+\text{H}]^+$  and  $[\text{NAK}+\text{H}]^+$  obtenues au niveau de calcul CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

		$\Delta E_n^\circ$ (kJ mol <sup>-1</sup> )		$\Delta H_n^\circ$ (kJ mol <sup>-1</sup> )			$\Delta S_n^\circ$ (J K <sup>-1</sup> mol <sup>-1</sup> )			$\Delta G_n^\circ$ (kJ mol <sup>-1</sup> )	
		msc <sup>a</sup>	average <sup>b</sup>	msc <sup>a</sup>	average <sup>b</sup>	exp[12]	msc <sup>a</sup>	average <sup>b,c</sup>	exp[12]	msc <sup>a</sup>	average <sup>b,c</sup>
[NAR + H] <sup>+</sup>	G <sub>1</sub>	-55.8	-52.7	-46.3	-43.8		-143.1	-129.8		-3.7	-5.1
	G <sub>2</sub>	-44.7	-46.9	-39.5	-40.5		-105.3	-110.7		-8.1	-7.6
	G <sub>3</sub>	-41.1	-41.8	-33.5	-34.5		-111.1	-110.3		-0.4	-1.6
[NAK + H] <sup>+</sup>	G <sub>1</sub>	-54.4	-55.7	-47.5	-48.7	-44.4±1.3	-107.4	-112.0	-73.6±4.2	-15.5	-15.4
	G <sub>2</sub>	-43.5	-45.4	-37.0	-38.5	-35.1±1.3	-102.6	-106.7	-61.1±4.2	-6.4	-6.5
	G <sub>3</sub>	-48.2	-46.6	-40.8	-39.3	-34.7±1.3	-119.4	-117.4	-69.9±4.2	-5.3	-4.5
[NAH + H] <sup>+</sup>	G <sub>1</sub>	-56.6	-56.2	-49.3	-49.1		-110.6	-109.9		-16.4	-16.4
	G <sub>2</sub>	-48.2	-44.6	-41.8	-37.9		-117.6	-97.6		-6.7	-8.8
	G <sub>3</sub>	-41.7	-43.3	-34.7	-36.8		-103.0	-111.9		-4.0	-3.5

<sup>a</sup> calcul effectué uniquement à partir du conformère le plus stable (Most Stable Conformer, msc) de chaque génération ; <sup>b</sup> calcul moyenné à partir des populations indiquées dans le tableau III.1.5. <sup>c</sup> L'entropie de mélange est incluse dans le calcul.

Plusieurs tendances peuvent être observées dans nos résultats. En général, l'enthalpie libre d'hydratation diminue avec l'augmentation du nombre de molécules d'eau, et l'on observe ainsi que la troisième molécule d'eau n'est liée que très faiblement aux ions étudiés, avec des valeurs inférieures à 5 kJ mol<sup>-1</sup> à 298 K. Cette faible interaction est due à des effets entropiques, puisque en termes énergétique ou enthalpique, les valeurs d'hydratation sont situées entre 35 et 55 kJ mol<sup>-1</sup>. De très faibles écarts sont observés entre les valeurs moyennes et les valeurs basées uniquement sur les conformères les plus stables, indiquant que soit le conformère le plus stable est majoritaire, soit les données d'hydratation des différents conformères sont peu différentes.

Nos données peuvent être comparées à celles de la littérature. En accord avec la littérature, qui indique que le groupe guanidinium interagit moins fortement qu'un groupe ammonium,[4] nous obtenons une enthalpie libre hydratation plus faible de  $[\text{NAR} + \text{H}]^+$  par rapport à  $[\text{NAK} + \text{H}]^+$ . Le terme entropique semble le responsable principal de cette différence. Néanmoins, on constate également une valeur légèrement inférieure pour les enthalpies, en particulier pour la valeur moyenne. Ce dernier aspect pourrait être mis en relation avec le fait que pour la génération G<sub>1</sub>, seul  $[\text{NAR} + \text{H}]^+(\text{H}_2\text{O})$  n'a pas 100% de sa population où la molécule d'eau est liée au groupe chargé. Pour la lysine N $\alpha$ -acétylée, les termes entropiques sont fortement surestimés par rapport aux données expérimentales, même si les variations d'une génération à l'autre sont approximativement reproduites. Pour le même système, les enthalpies sont très bien reproduites, avec une surestimation systématique de 3.5-4.5 kJ mol<sup>-1</sup> (en valeur absolue). En particulier, l'écart de 10 kJ mol<sup>-1</sup> entre la première et la seconde enthalpie d'hydratation, puis l'absence d'écart entre la seconde et la troisième, sont très bien reproduites.

Nos structures calculées permettent d'avancer une hypothèse pour expliquer ces variations : la première hydratation se fait systématiquement sur le site chargé, ce qui diminue probablement l'affinité du groupe ammonium pour une seconde molécule d'eau. De plus, pour l'ajout des deux molécules d'eau suivantes, il y a une compétition entre différents sites (ammonium, COOH, seconde sphère de solvation), indiquant que ces sites ont une affinité voisine et donc inférieure à la première affinité. Enfin, nous avons vu que la fixation d'une molécule d'eau sur le groupe COOH ne perturbe pas l'ammonium, suggérant que la fixation de la seconde molécule d'eau sur un des sites ne modifie que très faiblement l'affinité d'une troisième molécule d'eau pour les autres sites.

Cette discussion montre l'intérêt de connaître les données thermochimiques d'hydratation pour chaque site, individuellement. L'identification des filiations entre les structures permet de faire ces estimations en s'abstrayant des problèmes de changement de conformation (calcul des données dites « same conformation » (sc), voir détails dans le manuscrit ci-avant). Ainsi, les énergies et enthalpies d'hydratation des sites chargés sont rassemblées dans le tableau III.1.8.

Tableau III.1.8. Influence des liaisons hydrogène sur les énergies et enthalpies de première hydratation des sites chargés (calculs CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p)).

parent	Fils	Ratio	$\Delta E^{\circ}_{1,sc}$ (kJ.mol <sup>-1</sup> )	$\Delta H^{\circ}_{1,sc}$ (kJ.mol <sup>-1</sup> )
<b>[NAR+H]<sup>+</sup></b>				
<b>R<sup>0</sup>_1</b>	<b>R<sup>1</sup>_1</b>	0.42	-55.8	-46.3
<b>R<sup>0</sup>_4</b>	<b>R<sup>1</sup>_3</b>	0.15	-56.1	-49.1
<b>R<sup>0</sup>_3</b>	<b>R<sup>1</sup>_4</b>	0.08	-55.8	-47.3
<b>R<sup>0</sup>_8</b>	<b>R<sup>1</sup>_5</b>	0.07	-56.1	-47.9
<b>R<sup>0</sup>_5</b>	<b>R<sup>1</sup>_6</b>	0.05	-56.0	-47.5
<b>R<sup>0</sup>_9</b>	<b>R<sup>1</sup>_7</b>	0.04	-56.2	-49.6
<b>R<sup>0</sup>_4</b>	<b>R<sup>1</sup>_8</b>	0.02	-55.2	-46.5
	Moyenne		-55.9	-47.7
<b>M<sup>0</sup>_1</b>	<b>M<sup>1</sup>_1</b>		-67.6	-60.2
<b>M<sup>0</sup>_2</b>	<b>M<sup>1</sup>_2</b>		-56.7	-48.8
<b>[NAK+H]<sup>+</sup></b>				
<b>K<sup>0</sup>_1</b>	<b>K<sup>1</sup>_1</b>	0.35	-54.4	-47.5
<b>K<sup>0</sup>_1</b>	<b>K<sup>1</sup>_2</b>	0.28	-54.8	-47.9
<b>K<sup>0</sup>_3</b>	<b>K<sup>1</sup>_3</b>	0.24	-53.8	-46.3
<b>K<sup>0</sup>_2</b>	<b>K<sup>1</sup>_4</b>	0.06	-55.9	-48.4
<b>K<sup>0</sup>_7</b>	<b>K<sup>1</sup>_5</b>	0.04	-53.9	-46.9
<b>K<sup>0</sup>_3</b>	<b>K<sup>1</sup>_6</b>	0.03	-53.9	-45.6
	moyenne		-54.5	-47.1
<b>M<sup>0</sup>_3</b>	<b>M<sup>1</sup>_3<sup>a,b</sup></b>		-72.4	-66.4
<b>M<sup>0</sup>_4</b>	<b>M<sup>1</sup>_4</b>		-65.4	-57.8
<b>M<sup>0</sup>_5</b>	<b>M<sup>1</sup>_5</b>		-59.4	-52.0
<b>M<sup>0</sup>_6</b>	<b>M<sup>1</sup>_6</b>		-56.2	-49.1
<b>[NAH+H]<sup>+</sup></b>				
<b>H<sup>0</sup>_1</b>	<b>H<sup>1</sup>_1</b>	0.76	-56.6	-49.3
<b>H<sup>0</sup>_2</b>	<b>H<sup>1</sup>_2</b>	0.10	-55.9	-49.1
<b>H<sup>0</sup>_3</b>	<b>H<sup>1</sup>_3</b>	0.9	-54.5	-48.3
<b>H<sup>0</sup>_4</b>	<b>H<sup>1</sup>_4</b>	0.5	-56.5	-49.9
	moyenne		-55.9	-49.1
<b>M<sup>0</sup>_7</b>	<b>M<sup>1</sup>_7<sup>c</sup></b>		-65.4	-59.6
<b>M<sup>0</sup>_8</b>	<b>M<sup>1</sup>_8</b>		-55.8	-48.8

Données expérimentales disponibles : <sup>a</sup>  $\Delta H^{\circ}_1(\mathbf{M}^1_3) = -62.3 \pm 4.6, [6] -70.3, [7] -78.7 [8] \text{ kJ mol}^{-1}$ . <sup>b</sup>  $\Delta H^{\circ}_1(\text{n-C}_5\text{H}_{11}\text{NH}_3^+) = -61.9 \pm 0.8 \text{ kJ mol}^{-1}, [6]$  <sup>c</sup>  $\Delta H^{\circ}_1(\mathbf{M}^1_7) = -61.9 \text{ kJ mol}^{-1}, [14]$

Il est remarquable de constater que pour les différentes conformations d'un même acide aminé, l'enthalpie d'hydratation sur le site chargé est approximativement toujours la même. De plus, cette enthalpie d'hydratation est égale pour les trois acides aminés considérés, autour de -48 kJ mol<sup>-1</sup>. L'étude de modèles permet de comprendre ces tendances (Figure III.1.13).

Les sites chargés isolés (**M<sup>1</sup>\_1**, **M<sup>1</sup>\_3** et **M<sup>1</sup>\_7**) ont des énergies d'hydratation différentes : l'ammonium a une plus forte affinité ( $-66.4 \text{ kJ mol}^{-1}$ ) que le guanidinium ( $-60.2 \text{ kJ mol}^{-1}$ ) et l'imidazolium ( $-59.6 \text{ kJ mol}^{-1}$ ), vraisemblablement car sa charge positive n'est pas délocalisée. L'ajout d'une liaison hydrogène sur ces sites (**M<sup>1</sup>\_2**, **M<sup>1</sup>\_6** et **M<sup>1</sup>\_8**) diminue sensiblement leur affinité (de  $10 \text{ kJ mol}^{-1}$  pour le guanidinium et l'imidazolium et de  $15 \text{ kJ mol}^{-1}$  pour l'ammonium), conduisant pour les trois cas à une même valeur d'enthalpie d'hydratation d'environ  $-49 \text{ kJ mol}^{-1}$ . Ainsi, la présence d'une liaison hydrogène sur le site chargé explique que l'enthalpie d'hydratation sur le groupe NH du groupe chargé soit la même non seulement pour chaque conformère d'un même système, mais aussi pour les trois systèmes  $[\text{NAH} + \text{H}]^+$ ,  $[\text{NAK} + \text{H}]^+$  et  $[\text{NAR} + \text{H}]^+$ .

Par comparaison, nous avons évalué l'influence d'autres groupes que le CO d'amide du N-méthyl acétamide sur l'enthalpie d'hydratation d'un groupe ammonium (**M<sup>1</sup>\_4** et **M<sup>1</sup>\_5**). Un groupe COOH, moins donneur qu'un amide du fait de la plus forte électronégativité de l'oxygène par rapport à celle de l'azote, diminue plus faiblement l'affinité pour  $\text{H}_2\text{O}$ . De même dans le cas d'une liaison hydrogène contrainte. Ainsi, la force de l'interaction entre le site chargé et la molécule d'eau s'accroît lorsque celle entre le site chargé et l'accepteur de liaison hydrogène décroît, mettant en évidence la coopérativité de ces effets.

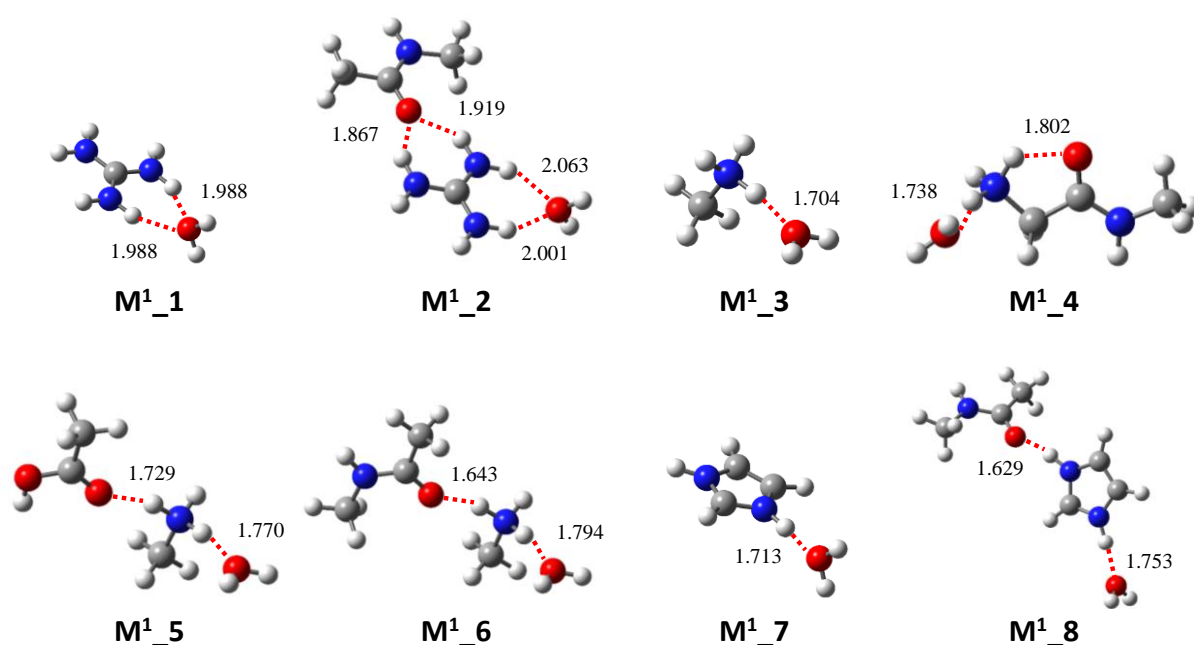


Figure III.1.13. Modèles pour la mono-hydratation des sites chargés guanidinium, ammonium et imidazolium.

De façon similaire, nous avons voulu estimer l'influence du site chargé et de sa nature sur la force de l'interaction entre le groupe acide et une molécule d'eau (Tableau III.1.9 et Figure III.1.14).

Tableau III.1.9. Influence du site chargé sur l'interaction  $\text{COOH}\cdots\text{OH}_2$ .

parent	Fils	M06/6-311++G(d,p)		ONIOM <sup>a</sup>	CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p)	
		$\Delta E_{1,sc}^\circ$ (kJ.mol <sup>-1</sup> )	$\Delta H_{1,sc}^\circ$ (kJ.mol <sup>-1</sup> )	$\Delta E_{1,sc}^\circ$ (kJ.mol <sup>-1</sup> )	$\Delta E_{1,sc}^\circ$ (kJ.mol <sup>-1</sup> )	$\Delta H_{1,sc}^\circ$ (kJ.mol <sup>-1</sup> )
<b>R<sup>0</sup>_1</b>	<b>R<sup>1</sup>_2</b>	-59.7	-52.6	/	-45.6	-38.5
<b>M<sup>0</sup>_9</b>	<b>M<sup>1</sup>_9</b>			-48.8		
<b>M<sup>0</sup>_10</b>	<b>M<sup>1</sup>_10</b>			-50.5		
<b>M<sup>0</sup>_11</b>	<b>M<sup>1</sup>_11</b>			-60.6		
<b>K<sup>0</sup>_1</b>	<b>K<sup>1</sup>_7</b>	-59.7	-53.9	/	-46.3	-40.4
<b>M<sup>0</sup>_12</b>	<b>M<sup>1</sup>_12</b>			-48.2		
<b>M<sup>0</sup>_13</b>	<b>M<sup>1</sup>_13</b>			-50.6		
<b>M<sup>0</sup>_14</b>	<b>M<sup>1</sup>_14</b>			-59.4		
<b>H<sup>0</sup>_1</b>	<b>H<sup>1</sup>_5</b>	-63.3	-56.4	/	-49.8	-43.0
<b>M<sup>0</sup>_15</b>	<b>M<sup>1</sup>_15</b>	-49.8	-42.4	/	-36.0	-28.6

<sup>a</sup> calculs au niveau ONIOM(M06/6-311++G(d,p):UFF)//M06/6-311++G(d,p) en ne considérant que l'énergie de la partie haute (calculée en DFT). Le découpage entre la partie haute et la partie basse est indiquée dans la Figure III.1.14.

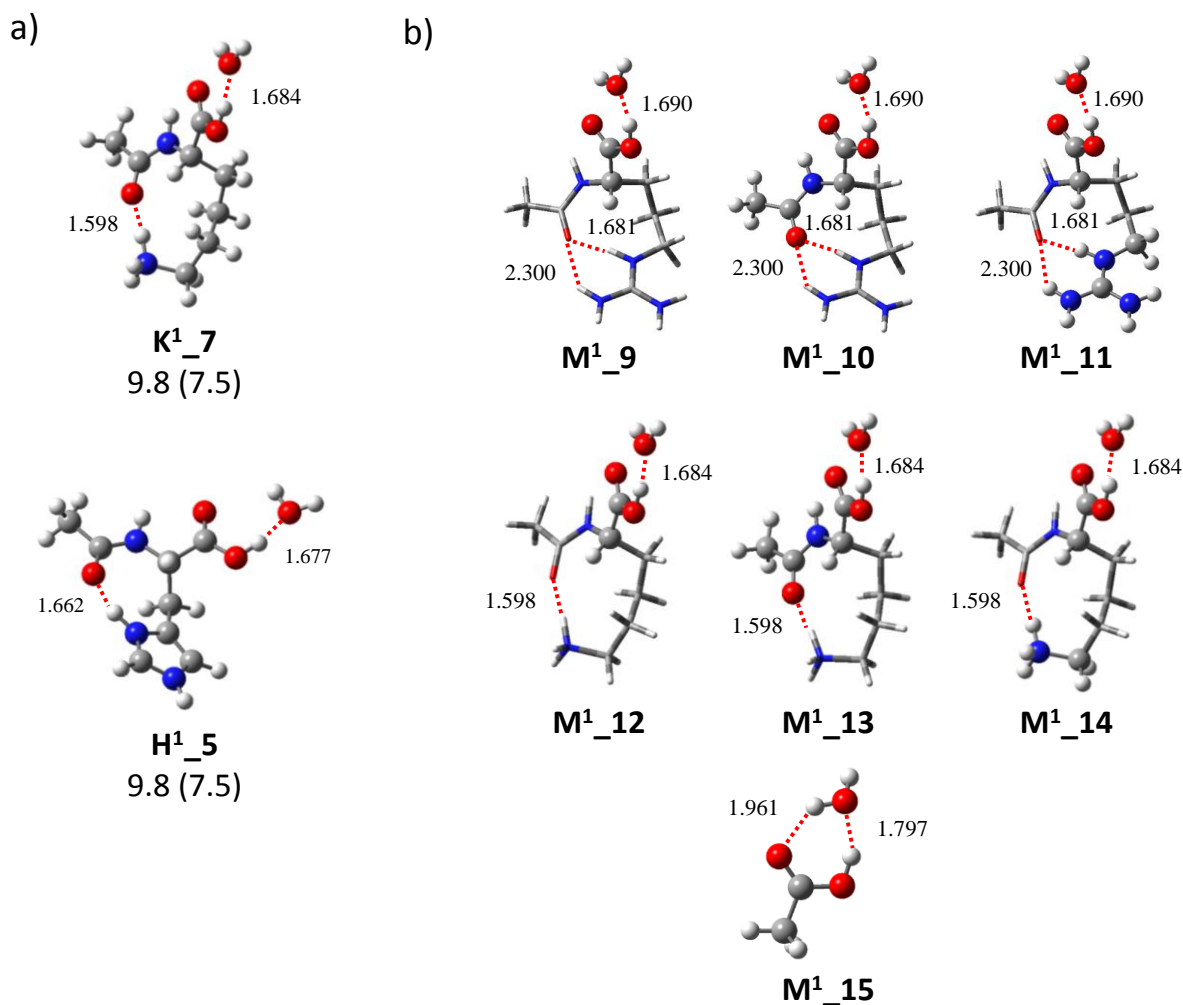


Figure III.1.14. Systèmes a) et b) utilisés dans le Tableau III.1.9. b) Pour les calculs ONIOM, les atomes inclus dans le niveau haut sont représentés en “ball and stick” tandis que ceux inclus dans le niveau bas sont en représentation “tube”. Enthalpies libres (et enthalpies entre parenthèses) relatives au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

Pour cela, nous avons réalisé des calculs en ne considérant que certains atomes de nos systèmes, en utilisant la méthode ONIOM. Ainsi, dans les modèles **M<sup>1</sup>\_12**, **M<sup>1</sup>\_9** et **M<sup>1</sup>\_15**, nous avons estimé l'énergie d'hydratation du groupe COOH seul (environ  $-49 \text{ kJ mol}^{-1}$ ). La présence du groupe amide voisin (modèles **M<sup>1</sup>\_10** et **M<sup>1</sup>\_13**) ne modifie que très faiblement cette affinité. A l'inverse, la présence d'un groupe chargé distant (modèles **M<sup>1</sup>\_11** et **M<sup>1</sup>\_14**) augmente fortement l'énergie d'hydratation, vraisemblablement du fait d'interactions électrostatiques entre la molécule de H<sub>2</sub>O et la charge. Ce résultat explique parfaitement l'énergie d'hydratation calculée pour les systèmes complets. La nature du groupe chargé ne semble pas jouer de rôle majeur, même si dans le cas de l'histidine, on constate une affinité légèrement plus forte.



Enfin, la même procédure consistant à déterminer les valeurs d'enthalpies d'hydratation pour des conformères identiques (« sc ») a été appliquée pour l'addition de la seconde et troisième molécule d'eau, en fonction du site de fixation de celles-ci (Tableau III.1.10).

Tableau III.1.10. Valeur moyenne des secondes ( $\Delta H^\circ_{2,sc}$ ) et troisièmes ( $\Delta H^\circ_{3,sc}$ ) enthalpies d'hydratation en fonction du site de fixation de la molécule d'eau au niveau CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p).

	Site chargé	$\Delta H^\circ_{2,sc}$ (kJ.mol <sup>-1</sup> )		$\Delta H^\circ_{3,sc}$ (kJ.mol <sup>-1</sup> )	
		Seconde sphère de solvation	COOH	Seconde sphere de solvation	COOH
[NAR + H] <sup>+</sup>	-42.2 <sup>a</sup> -32.9 <sup>b</sup>	-35.6	-39.5	-34.1	/
[NAK + H] <sup>+</sup>	-40.7	-37.3	-40.3	-33.2	-38.8
[NAH + H] <sup>+</sup>	/	-35.6	-41.8	-34.8	/

<sup>a</sup> molécule d'eau en pont entre deux groupes NH du guanidinium. <sup>b</sup> molécule d'eau liée à un seul groupe NH du guanidinium

La comparaison avec les données du tableau III.1.8 montre une diminution de l'enthalpie d'hydratation sur le site chargé. La première molécule d'eau étant placée sur ce même site, cette deuxième liaison hydrogène induit une nouvelle diminution de l'affinité du site chargé pour une deuxième molécule d'eau. Inversement, l'affinité pour le groupe COOH n'est que très faiblement affectée par la présence de la première molécule sur le groupe chargé (comparaison avec le tableau III.1.9), voire de la deuxième dans le cas de [NAK + H]<sup>+</sup>, confirmant ainsi la non influence entre ces deux sites. Ces deux évolutions conduisent à des valeurs d'enthalpie d'hydratation proches, ce qui explique la compétition observée pour ces deux sites dans les structures de la génération **G**<sub>2</sub>. La fixation sur la première molécule d'eau, en seconde sphère de solvation, n'est que légèrement plus faible (et est légèrement plus favorable d'un point de vue entropique, Tableau III.1.5), expliquant que ce site est lui aussi compétitif.

Ces données expliquent l'évolution des enthalpies d'hydratation pour les trois systèmes [NAK + H]<sup>+</sup>, [NAR + H]<sup>+</sup> et [NAH + H]<sup>+</sup> (Tableau III.1.7) :

- [NAK + H]<sup>+</sup>. La première molécule d'eau se fixe sur l'ammonium avec une forte affinité (-48.7 kJ mol<sup>-1</sup>). Les molécules d'eau suivantes se fixent sur l'ammonium, de plus faible affinité, ou sur le COOH ou en deuxième sphère de solvation, tous sites

d'affinité comparables, entraînant des seconde et troisième enthalpies d'hydratation plus faible que la première mais équivalentes entre elles (-38.5 et -39.3 kJ mol<sup>-1</sup>).

- [NAH + H]<sup>+</sup>. La première molécule d'eau se fixe sur l'imidazolium avec une forte affinité (-49.1 kJ mol<sup>-1</sup>). La seconde molécule d'eau se fixe essentiellement sur le COOH ou en deuxième sphère de solvation, avec donc une seconde enthalpie d'hydratation plus faible que la première (-37.9 kJ mol<sup>-1</sup>). La troisième molécule d'eau complète le site manquant, avec une enthalpie d'hydratation proche mais légèrement inférieure à la précédente (-36.8 kJ mol<sup>-1</sup>).
- [NAR + H]<sup>+</sup>. La première molécule d'eau se fixe sur le guanidinium, mais aussi minoritairement (17%) sur le groupe COOH, expliquant la plus faible enthalpie d'hydratation par rapport aux deux autres systèmes (-43.8 kJ mol<sup>-1</sup>). La seconde molécule d'eau complète l'un des deux sites précédents ou éventuellement se place en seconde sphère de solvation. En moyenne, cela donne une faible diminution de l'enthalpie d'hydratation (-40.5 kJ mol<sup>-1</sup>). La troisième molécule d'eau est essentiellement en seconde sphère de solvation, expliquant la plus faible valeur obtenue (-34.5 kJ mol<sup>-1</sup>).

Pour terminer cette partie, nous nous sommes interrogés sur la possibilité que les données thermodynamiques d'hydratation expérimentales puissent servir de sonde pour déterminer la structure locale autour d'un groupe chargé et le site de localisation de la molécule d'eau. Pour cela, nous avons considérés les données disponibles (Tableau III.1.1) pour la première enthalpie d'hydratation pour des peptides avec un résidu basique et N $\alpha$ -acétylés (Tableau III.1.11).

Tableau III.1.11. Extrapolation de nos résultats sur des peptides plus étendus.

	$\Delta H^{\circ}_{1,\text{exp}}$ (kJ mol <sup>-1</sup> )	Environnement local du site chargé	Site de la première hydratation
[Ac-AAKAA + H] <sup>+</sup>	-35.6[12]	Deux ou trois liaisons hydrogène	OH ou NH neutre, ou <u>insertion</u>
[Ac-AAAK + H] <sup>+</sup>	-28.9[12]	Deux ou trois liaisons hydrogène	OH ou NH neutre, ou <u>insertion</u>
[Ac-KA <sub>15</sub> + H] <sup>+</sup>	-48.9[3]	Une liaison hydrogène	Ammonium
[Ac-KA <sub>20</sub> + H] <sup>+</sup>	-47.0[3]	Une liaison hydrogène	Ammonium
[Ac-A <sub>15</sub> H + H] <sup>+</sup>	-62.7[3]	Zéro liaison hydrogène	Imidazolium
[Ac-AARAA + H] <sup>+</sup>	-39.7[4]	Deux liaisons hydrogène	COOH group

$[\text{Ac-AARAA} + \text{H}]^+$  possède une enthalpie d'hydratation proche de la première, et surtout de la deuxième de  $[\text{NAR} + \text{H}]^+$ . En conséquence, la présence de deux liaisons hydrogène autour du guanidinium (ou éventuellement d'une seule) est le plus probable.  $[\text{Ac-A}_{15}\text{H} + \text{H}]^+$  présente une forte enthalpie d'hydratation qui ne semble pouvoir s'expliquer que par l'absence de liaison hydrogène autour du groupe chargé, qui fixerait alors la première molécule d'eau. Les peptides  $[\text{Ac-KA}_{15} + \text{H}]^+$  et  $[\text{Ac-KA}_{20} + \text{H}]^+$  ont des premières enthalpies d'hydratation très proches de celle de  $[\text{NAK} + \text{H}]^+$ . Sur cette base, on prévoit que l'ammonium de ces deux peptides ne forme qu'une seule liaison hydrogène (ce qui ne correspond pas à la structure proposée sur la base de travaux réalisés à l'aide du champ de forces CHARMM[3]), et que la première molécule d'eau viendrait s'y attacher. Enfin, les peptides  $[\text{Ac-AAKAA} + \text{H}]^+$  et  $[\text{Ac-AAAAK} + \text{H}]^+$  ont des enthalpies d'hydratation faibles qui ne peuvent s'expliquer que par la présence de deux, voire trois, liaisons hydrogène autour de l'ammonium, en accord avec les structures proposées par les auteurs de l'étude expérimentale suite à des travaux de modélisation moléculaire avec le champ de forces AMBER.[12]

Cette analyse rapide montre la possibilité d'obtenir des informations structurales à partir de données thermodynamiques d'hydratation. Cependant, celles-ci restent pour l'instant encore largement spéculatives et des études complémentaires seraient nécessaires pour les affermir.

### III.1.4. Conclusions

Dans cette partie, nous avons élaboré une nouvelle stratégie d'exploration de la surface d'énergie potentielle de molécules flexibles et appliquée celle-ci à l'étude des  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$ . Notre stratégie repose sur (i) la combinaison de l'approche hiérarchique avec l'approche arbre de Darwin qui permettent de générer environ 60 % et 40 % des structures  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$ , respectivement et (ii) l'utilisation successive du champ de forces AMOEBA, de la méthode semi-empirique PM6-DH2 et de la DFT (M06/6-311++G(d,p)), avec CP-MP2/6-311++G(2d,2p) comme niveau pour calculer les énergies finales. Cette stratégie nous a permis d'obtenir une population de conformères la plus complète possible dans la gamme 0-7.5 kJ mol<sup>-1</sup> en enthalpie libre. Les données thermochimiques calculées au niveau de théorie CP-MP2/6-311++G(2d,2p)//M06/6-311++G(d,p) sont en accord avec les données expérimentales disponibles dans la littérature, validant ainsi notre stratégie de génération des structures  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})$ .

Les structures les plus stables des systèmes  $[NAX + H]^+(H_2O)_n$  possèdent une liaison hydrogène intramoléculaire, généralement de type  $CO_{amide} \cdots HN^+$ . En hydratant progressivement ces systèmes, la première molécule d'eau se place sur le site chargé, sans modification de la conformation du squelette peptidique. Compte tenu de la disponibilité des liaisons NH des sites chargés, la deuxième molécule d'eau se place, généralement, de nouveau sur le site chargé pour  $[NAK + H]^+(H_2O)_2$  et sur l'acide carboxylique pour  $[NAH + H]^+(H_2O)_2$  et  $[NAR + H]^+(H_2O)_2$ . La troisième se fixe le plus souvent en deuxième sphère de solvation en interagissant avec la première molécule d'eau pour  $[NAH + H]^+(H_2O)_2$  et  $[NAR + H]^+(H_2O)_2$ , tandis qu'elle va se placer sur l'acide carboxylique dans le cas  $[NAK + H]^+(H_2O)_3$ . Les premières structures dans lesquels une molécule d'eau perturbe la conformation du squelette peptidique, en s'insérant dans une liaison hydrogène intramoléculaire ou en formant un pont de liaisons hydrogène entre deux sites distants, commencent à apparaître à  $n = 2$  pour  $[NAH + H]^+(H_2O)_n$  et à  $n = 3$  pour les deux autres systèmes. Elles ne sont toutefois pas majoritaires dans la population de conformères et on peut émettre l'hypothèse qu'elles devraient être plus importantes à partir de  $n = 4$ .

Les énergies de fixation des molécules d'eau dans les différents sites dépendent à la fois des groupements dans le squelette peptidique à proximité du site de fixation (site chargé, COOH, liaison hydrogène intramoléculaire) mais aussi de la présence d'autres molécules d'eau. Nous avons choisi de nommer ce phénomène « interaction réciproque ». De plus, nous avons mis en évidence que les données expérimentales ne peuvent systématiquement permettre de distinguer les sites préférentiels d'hydratation car plusieurs structures avec des sites d'hydratation différents peuvent être présentes en quantité non négligeable. Enfin, en combinant les données thermochimiques expérimentales avec l'ordre préférentiel des sites d'hydratation établi théoriquement, nous avons pu supposer une structuration locale des sites chargés, solvatés ou non par des groupements carbonyles, pour des peptides de plus grandes tailles comme des pentapeptides.

Encouragés par ces travaux portant sur la micro-hydratation d'acides aminés chargés où nous avons mis au point une stratégie de modélisation efficace, nous avons poursuivi notre étude par celle de clusters de glycine bêtaïne. La taille des systèmes cibles nécessitait, là aussi, comme nous le détaillons dans la suite de ce chapitre, une mise en place précise d'une stratégie de modélisation pour déterminer les structures les plus pertinentes à analyser.

### III.1.5. Bibliographie

- [1] Addario, V.; Guo, Y.; Chu, I. V.; Ling, Y.; Ruggerio, G.; Rodriguez, C. F.; Hopkinson, A. C.; Siu, K. W. M. *Int. J. Mass Spectrom.* **2002**, *219*, 101-114.
- [2] Bouchoux, G. *Mass Spectrom. Rev.* 2014, doi: 10.1002/mas.21416.
- [3] Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2002**, *124*, 11148-11158.
- [4] Liu, D.; Wyttenbach, T.; Barran, P. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2003**, *125*, 8458-8464.
- [5] Wyttenbach, T.; Bowers, M. T. *Chem. Phys. Lett.* **2009**, *480*, 1-16.
- [6] Gilligan, J. J.; Lampe, F. W.; Nguyen, V. Q.; Vieira, N. E.; Yergey, A. L. *J. Phys. Chem. A* **2003**, *107*, 3687-3691.
- [7] Meot-Ner, M. *J. Am. Chem. Soc.* **1984**, *106*, 1265-1272.
- [8] Lau, Y. K.; Kebarle, P. *Can. J. Chem.* **1981**, *59*, 151-155.
- [9] Wincel, H. *Chem. Phys. Lett.* **2007**, *439*, 157-161.
- [10] Klassen, J. S.; Blades, A. T.; Kebarle, P. *J. Phys. Chem.* **1995**, *99*, 15509-15517.
- [11] Gao, B.; Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2009**, *113*, 9995-10000.
- [12] Liu, D.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass Spectrom.* **2004**, *236*, 81-90.
- [13] Wyttenbach, T.; Bowers, M. T. *Top. Curr. Chem.* **2003**, *225*, 207-232.
- [14] Meot-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3075-3080.
- [15] Jacquemin, D.; Michaux, C.; Perpète, E. A.; Frison, G. *J. Phys. Chem. B* **2011**, *115*, 3604-3613.
- [16] Fischer, G. *Chem. Soc. Rev.* **2000**, *29*, 119-127.
- [17] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.
- [18] Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344-352.
- [19] Korth, M. *J. Chem. Theory Comput.* **2010**, *6*, 3808-3816.
- [20] Michaux, C.; Wouters, J.; Jacquemin, D.; Perpète, E. A. *Chem. Phys. Lett.* **2007**, *445*, 57-61.
- [21] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 2430-2438.
- [22] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 9896-9902.
- [23] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 7702-7705.
- [24] Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 632-638.
- [25] Campo-Cacharrón, A.; Cabaleiro-Lago, E. M.; Rodríguez-Otero, J. *Theor. Chem. Acc.* **2012**, *131*, 1290-1302.

## Partie II : Clusters de glycine bétaine en phase gazeuse



## III.2.1. Introduction

### III.2.1.1. Définition de la glycine bêtaïne

La principale différence qui existe entre les acides aminés en phase gazeuse et en solution est que les formes zwitterionique et canonique prédominent en solution et en phase gazeuse, respectivement. La Figure III.2.1 présente la glycine sous ses deux formes neutres a) et b). La glycine bêtaïne, notée GB, est un dérivé de la glycine zwitterionique dont les atomes d'hydrogène portés par l'ammonium ont été substitués par des groupements méthyles c). Cette molécule possède donc une charge positive fixe permanente et, lorsqu'elle est globalement neutre, un groupement  $\text{COO}^-$  anionique. Il s'agit donc d'une molécule particulièrement intéressante en tant que modèle de la forme zwitterionique de la glycine. De plus, le groupement carboxylate peut être protoné d). La glycine bêtaïne protonée sera par la suite noté  $[\text{GB} + \text{H}]^+$ .

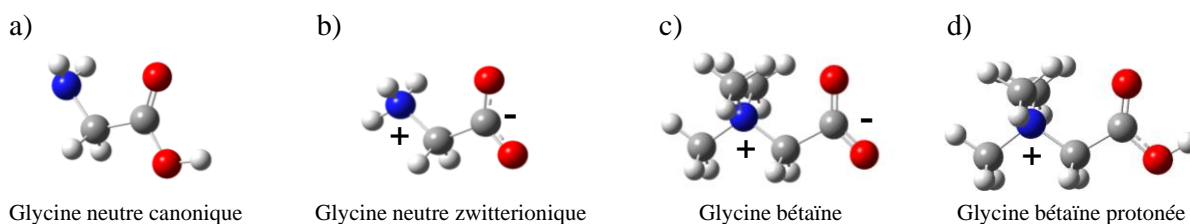


Figure III.2.1. Différentes formes de la glycine.

Dans ce qui suit, les clusters de glycine bêtaïne multiprotonés formés par  $(n-z)\text{GB}$  et par  $z[\text{GB} + \text{H}]^+$  seront notés  $[\text{GB}_n + z\text{H}]^{z+}$ .

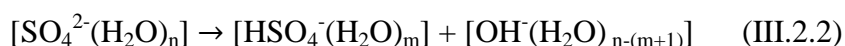
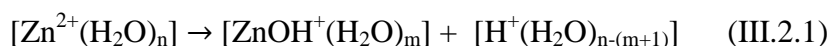
### III.2.1.2. Travaux antérieurs

#### III.2.1.2.1. Exemple de Clusters étudiés dans la littérature

Les clusters sont des phases transitoires entre la phase gazeuse et la phase condensée. Par exemple, en étudiant des tailles croissantes de clusters, il est possible d'étudier graduellement la transition entre les propriétés en phase gazeuse et celles en phase condensée. En spectrométrie de masse, l'étude des clusters ioniques est devenue un sujet d'étude majeur qui se concentre principalement sur leurs formations, leurs structures et leurs dissociations par voie vibrationnelle (CID) et électronique (EID). Nous avons étudié un premier exemple de clusters  $[\text{NAX} + \text{H}]^+(\text{H}_2\text{O})_n$  avec  $\text{X} = \text{His}, \text{Arg}$  et  $\text{Lys}$  dans la première partie de ce chapitre.

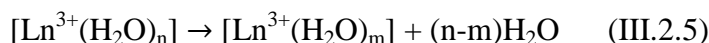
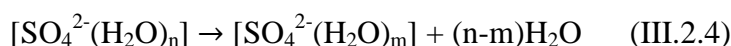
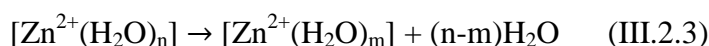


D'autres clusters issus de la complexation entre  $n$  molécules d'eau et un ion de charge  $z^+$  ou  $z^-$  ont fait l'objet de travaux en spectrométrie de masse, comme par exemple les clusters de type  $[M^{z+/-}(H_2O)_n]$  avec  $M^{z+/-} = Zn^{2+}, [1-3] SO_4^{2-}, [4,5] Ln^{3+}, [6,7] Pb^{2+} [8]$ . Lors de l'étude CID sur ces systèmes, deux processus de fragmentation sont en compétition. Le premier, noté (i), est la réduction de la charge par l'évaporation d'un ion, ce qui implique une séparation de charge du cluster initial:



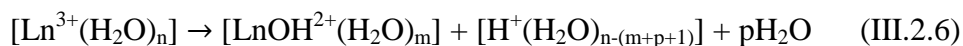
Dans ces deux exemples, il reste  $m$  molécules de  $H_2O$  autour de  $Zn^{2+}$  et  $SO_4^{2-}$  et  $n-(m+1)$  molécules de  $H_2O$  environnent l'ion qui s'est évaporé,  $H^+$  et  $OH^-$ , respectivement.

L'autre processus de dissociation, noté (ii), correspond à l'évaporation d'un neutre :



Dans ces trois exemples,  $n-m$  molécules de  $H_2O$  se sont évaporées et il reste  $m$  molécules de  $H_2O$  autour de l'ion.

Les deux processus (i) et (ii) peuvent également avoir lieu en même temps ; dans l'exemple qui suit, il y a eu  $n-(m+p+1)$  molécules de  $H_2O$  qui se sont évaporées avec l'ion  $H^+$  et  $p$  molécules de  $H_2O$  qui se sont évaporées en plus.



L'augmentation de la taille des clusters  $[SO_4^{2-}(H_2O)_n]$  et  $[Zn^{2+}(H_2O)_n]$  induit un changement de voie de fragmentation allant du processus de la séparation de charge vers l'élimination d'une molécule de  $H_2O$  à partir de  $n = 6$  et  $7$  pour  $[SO_4^{2-}(H_2O)_n]$  et  $[Zn^{2+}(H_2O)_n]$  respectivement. Les clusters  $[La^{3+}(H_2O)_n]$  tel que  $n = 17$  à  $28$  se dissocient suivant les voies (i) et (ii) qui sont en compétition.[6]

Il existe d'autres catégories de clusters ; comme par exemple des homo-[9-18] et hétéro-[11,15,18] dimères d'acides aminés, de peptides monoprotoné[19,20] et des clusters de

peptides multichargés.[21,22] Ces types de systèmes permettent de mieux comprendre (i) les interactions existantes entre plusieurs résidus de peptides et protéines, et (ii) les processus de transfert de proton intervenant dans la catalyse enzymatique, les reconnaissances moléculaires notamment. Les propriétés thermochimiques, comme la basicité en phase gazeuse et les énergies de liaisons,[15,23] et les propriétés structurales, obtenues par IRPD et IRMPD,[9-13] y sont étudiées. Avec l'enregistrement des spectres IRPD et IRMPD, l'un des objectifs est de déterminer la forme adoptée (zwitterionique ou canonique) par l'acide aminé ne portant pas le proton. Il semblerait que la forme canonique soit privilégiée dans les homo et hétéro- dimères protonés avec des chaînes aliphatiques[9,11], tandis que la forme zwitterionique l'est pour  $[\text{Pro}_2+\text{H}]^+$ . [9] Les conclusions sont plus difficiles à obtenir avec  $[\text{Lys}_2+\text{H}]^+$ . En effet, la forme zwitterionique permet d'expliquer la zone des  $2700\text{-}3700\text{ cm}^{-1}$  [13] du spectre IRPD, alors que celle à  $1000\text{-}1900\text{ cm}^{-1}$  correspond plutôt à la forme canonique.[12,13] Kong a donc supposé que les deux formes coexistent mais que les systèmes formés durant l'ionisation sont sensibles aux conditions expérimentales.

Les clusters multichargés de leucine-encéphaline ( $[\text{LE}_n]^{z+}$ ) ont des voies de fragmentation similaire à celle des clusters du type  $[\text{M}^{z+/-}(\text{H}_2\text{O})_n]$ . Les voies de fragmentations correspondant à une séparation de charge et à l'élimination d'un neutre dépendent de la valeur du terme  $\frac{z^2}{n}$ , avec  $z$  la charge totale du cluster et  $n$  le nombre de monomères de leucine-encéphaline. Pour des faibles (hautes) valeurs en  $\frac{z^2}{n}$ , i.e.  $n$  grand (petit), l'évaporation d'un neutre (la séparation de charge) est privilégiée.[22] Ces deux types de fragmentation sont en compétition pour des  $\frac{z^2}{n} \sim 0.5$ . La photo-fragmentation du dimère monoprotoné de la leucine-encéphaline[19,20] a été réalisée, contribuant à la compréhension des dommages créés par les radiations sur les biomolécules.

### III.2.1.2.2 Clusters de glycine bêtaïne dans la littérature

Les clusters de GB ont également été étudiés via la spectrométrie de masse,[24-27] les résultats obtenus sont présentés dans cette partie. L'énergie de dissociation de  $[\text{GB}_2 + \text{H}]^+$ , mesurée expérimentalement,[23] a été estimée à  $139.9 \pm 4.8\text{ kJ mol}^{-1}$ . O'Hair et son équipe ont étudié une solution de glycine bêtaïne en utilisant une source électrospray couplée avec un LTQ-FT ICR (trappe ionique linéaire, LTQ). Le spectre de masse obtenue est représenté à la Figure III.2.2. Il met en évidence l'existence de clusters mono ( $[\text{GB}_n + \text{H}]^+$  et  $[\text{GB}_n + \text{Na}]^+$ ) et multichargés  $[\text{GB}_n + z\text{H}]^{z+}$ . Les clusters du type  $[\text{GB}_n + 2\text{H}]^{2+}$  et  $[\text{GB}_n + 4\text{H}]^{4+}$  sont facilement

identifiable lorsque  $n$  est impair, et lorsque  $n$  n'est pas un multiple de 3 pour le cluster  $[\text{GB}_n + 3\text{H}]^{3+}$ . La haute résolution du LTQ-FT ICR permet de distinguer les clusters multichargés de même  $m/z$ , par exemple  $[\text{GB} + \text{H}]^+$  et  $[\text{GB}_2 + 2\text{H}]^{2+}$ , grâce aux pics isotopiques. Les plus petits clusters observés sont :  $[\text{GB}_2 + \text{H}]^+$ ,  $[\text{GB}_2 + \text{Na}]^+$ ,  $[\text{GB}_{13} + 2\text{H}]^{2+}$  et  $[\text{GB}_{25} + 3\text{H}]^{3+}$  ; donc la taille minimale du cluster dépend de la charge totale. Les clusters  $[\text{GB}_2 + \text{H}]^+$ ,  $[\text{GB}_2 + \text{Na}]^+$ ,  $[\text{GB}_{15} + 2\text{H}]^{2+}$  et  $[\text{GB}_{21} + 2\text{H}]^{2+}$  ont fait l'objet d'une étude détaillée par CID et ECD.[24-27] Le plus petit cluster doublement chargé  $[\text{GB}_{13} + 2\text{H}]^{2+}$  n'a pas été étudié du fait de sa faible intensité.

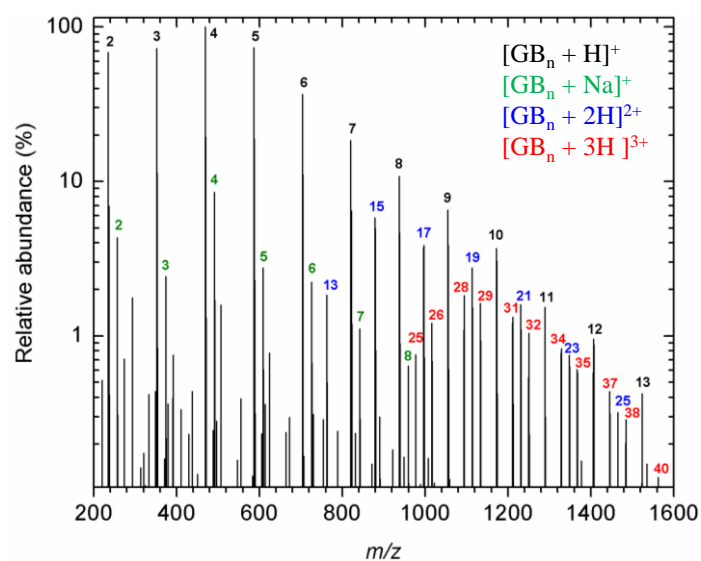
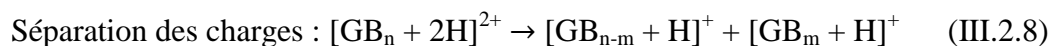
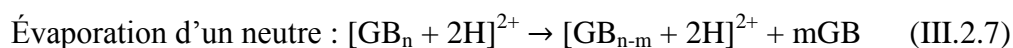


Figure III.2.2. Spectre de masse ESI/MS haute résolution obtenue avec un LTQ-FT ICR d'une solution de bétaine.[24]

Les spectres CID des clusters  $[\text{GB}_{15} + 2\text{H}]^{2+}$  et  $[\text{GB}_{21} + 2\text{H}]^{2+}$  issus des références [24] et [27] ont été reportés dans la Figure III.2.3. Les composés obtenus après fragmentation sont soit mono-chargés soit doublement chargés. Ainsi, deux voies de fragmentations ont été identifiées et elles sont similaires à celles décrites précédemment (voir III.2.2. Généralités) :



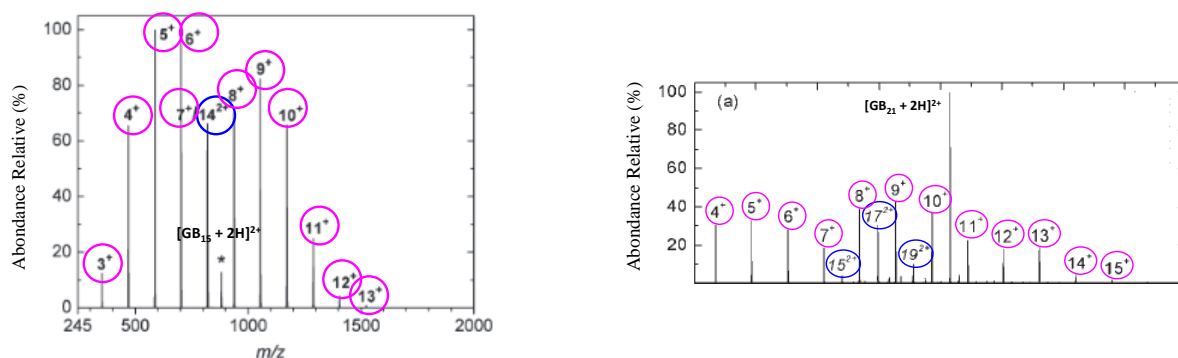
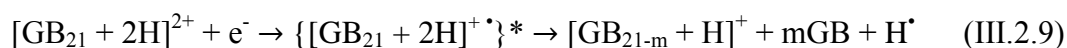
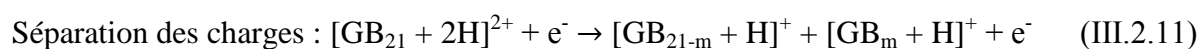
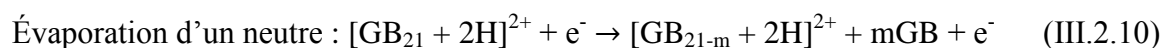


Figure III.2.3. Spectre de masse CID des clusters  $[\text{GB}_{15} + 2\text{H}]^{2+}$  à gauche[27] et  $[\text{GB}_{21} + 2\text{H}]^{2+}$  à droite[24]. Les clusters  $[\text{M}_n + \text{H}]^+$  et  $[\text{M}_n + 2\text{H}]^{2+}$  sont entourés en rose et en bleu respectivement.

Les spectres obtenus après EID dépendent fortement de l'énergie des électrons. Pour des faibles énergies allant jusqu'à 10 eV, l'espèce excitée  $\{[\text{GB}_{21} + 2\text{H}]^+\}^*$  formée se stabilise en éjectant  $\text{H}^\bullet$  et  $m$  molécules de GB :



Pour des énergies au-dessus de 12 eV, l'évaporation de  $m$  GB sans élimination d'un  $\text{H}^\bullet$  est observée ainsi que l'explosion coulombienne des clusters multichargés :



Dans cette même gamme d'énergie, il y a également l'apparition de clusters monochargés résultant de la rupture de la liaison de type C-X :  $[\text{GB}_m + \text{H} + (\text{GB}-\text{CO}_2)]^+$  et  $[\text{GB}_m + (\text{GB}-\text{CO}_2)]^+$  issus d'un processus de décarboxylation,  $[\text{GB}_m + \text{C}_3\text{H}_8\text{N}]^+$ ,  $[\text{GB}_m + \text{CH}_3]^+$  issus d'un transfert de  $\text{CH}_3$ .

Les mêmes voies de fragmentations ont été observés en EID pour le cluster  $[\text{GB}_{15} + 2\text{H}]^{2+}$ , excepté l'évaporation de  $m$  GB. Ainsi, que ce soit en CID et en EID, les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  se dissocient par élimination d'un ou plusieurs GB, ainsi que par séparation de charge que l'on nommera quelque fois « explosion coulombienne ».

Dans le cadre de notre étude, nous avons choisi d'étudier le système en focalisant nos efforts sur les processus d'évaporation d'un neutre et de séparation des charges dans  $[\text{GB}_n + 2\text{H}]^{2+}$ . Ces deux processus apparaissent en CID mais aussi en EID à partir d'une même gamme d'énergie des électrons. De plus, ils correspondent aux processus où seules des interactions non-covalentes sont rompues, contrairement aux autres, comme celui représenté dans l'équation III.2.9, qui impliquent des ruptures de liaisons covalentes. Ils nous permettront de connaître les principes de bases que donnent les différentes fragmentations observées et les caractéristiques fondamentales des clusters de GB. Des constatations expérimentales énoncées ci-dessus, plusieurs questions se posent :

- Quelle est la structure de ces clusters multichargés ?
- Où sont localisées les GB protonées au sein de ces clusters ?
- Pourquoi le plus petit cluster di-chargé observé expérimentalement est le  $[\text{GB}_{13} + 2\text{H}]^{2+}$  ?
- Quelle est l'énergie de cohésion, d'élimination d'une ou plusieurs GB et de l'explosion coulombienne au sein de ces clusters ?
- Quelle est la nature des interactions non-covalentes au sein de ces clusters ?

Pour répondre à ces questions, nous avons étudié les systèmes  $[\text{GB}_n + 2\text{H}]^{2+}$  avec  $n = 2$  à 6, 13.

## III.2.2. Stratégie d'étude

### *III.2.2.1. Stratégie d'exploration de la surface d'énergie potentielle*

La flexibilité et la complexité de ces clusters nécessitent l'élaboration d'une stratégie d'exploration de la surface d'énergie potentielle surtout pour identifier la structure la plus stable de  $[\text{GB}_{13} + 2\text{H}]^{2+}$ . Afin d'explorer la surface d'énergie potentielle des clusters de GB, nous avons généré les structures par dynamique moléculaire en utilisant le champ de force polarisable AMOEBA. La première étape consistait à obtenir les paramètres de GB et  $[\text{GB} + \text{H}]^+$ , qui n'étaient pas directement disponible dans le fichier de paramètres AMOEBA, en les extrayant de molécules déjà paramétrées. Les systèmes sélectionnés sont reportés dans le Tableau III.2.1. Nous avons choisi de tester directement ces paramètres pour l'exploration de

la surface d'énergie potentielle, sans ré-optimisation des paramètres de torsions et de multipôles atomiques de GB et  $[\text{GB} + \text{H}]^+$ .

Tableau III.2.1. Définition des atomes dans le fichier de paramètre AMOEBA.

type	classe	atomes	
236	30	C	C-Terminal $\text{COO}^-$
237	31	O	C-Terminal $\text{COO}^-$
2	2	CA	Glycine CA
6	6	H	Glycine HA
234	41	N	N-Terminal $\text{NH}_3^+$
244	16	C	N-Terminal PRO CD
245	6	H	N-Terminal PRO HD
604	30	C	Gly C-Terminal C=O
605	32	O	Gly C-Terminal O=C
606	33	O	Gly C-Terminal OH
607	34	H	Gly C-Terminal HO

Pour l'exploration de la surface d'énergie potentielle, nous avons construit différents clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ , présentant différents types d'interactions, comme point de départ des dynamiques moléculaires. Pour une même structure de départ, nous avons lancé en parallèle plusieurs dynamiques à des températures différentes (50, 100, 150 et 300 K) avec des temps de dynamique de 0.16 et 4.0 ns pour lesquels les structures ont été sauvegardées tous les 0.16 et 4.0 ps, respectivement. Dans un premier temps, toutes les structures générées lors de ces dynamiques moléculaires ont été optimisées avec AMOEBA et les minima obtenus ont été classés par famille en fonction de leurs interactions intermoléculaires. Les structures les plus stables en énergie de chaque famille ont été optimisées avec la méthode M06 associée à la base d'orbitale atomique 6-31+G(d,p). Cette première approche, réalisée sur les clusters  $[\text{GB}_3 + 2\text{H}]^{2+}$  et  $[\text{GB}_4 + 2\text{H}]^{2+}$ , nous a permis de tester la fiabilité de nos paramètres utilisés pour définir les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  avant d'étudier  $[\text{GB}_5 + 2\text{H}]^{2+}$  et  $[\text{GB}_{13} + 2\text{H}]^{2+}$ . Les résultats de cette exploration ont été reportés dans la Figure III.2.4.

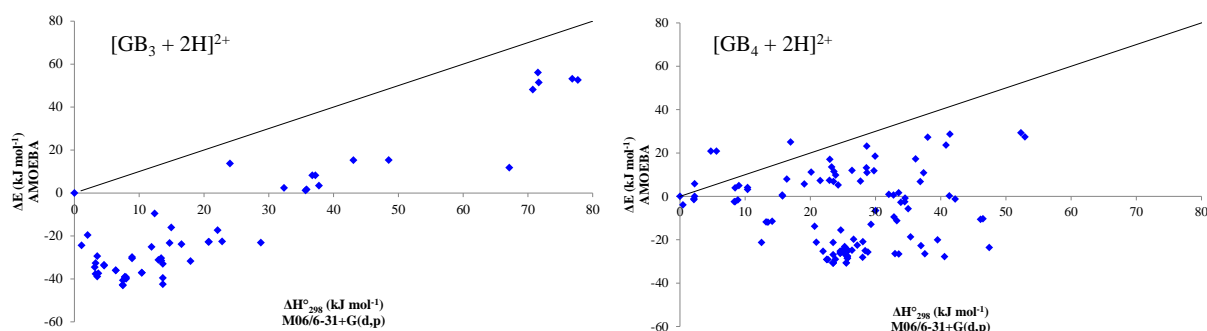


Figure III.2.4. Corrélation énergétique entre AMOEBA et M06/6-31+G(d,p) ( $\text{kJ mol}^{-1}$ ).

De ces données, nous avons constaté que les énergies AMOEBA ne sont pas bien corrélées aux enthalpies M06/6-31+G(d,p), ce qui est dû aux différences géométriques et/ou énergétiques. Par conséquent, nous avons calculé le Root Mean Square Deviation (RMSD) des géométries  $[\text{GB}_3 + 2\text{H}]^{2+}$  AMOEBA par rapport aux géométries M06/6-31+G(d,p). Les résultats obtenus sont reportés dans la Figure III.2.5. On constate que 15 structures ont un RMSD supérieur à 1,5 Å, ce qui est loin d'être satisfaisant.

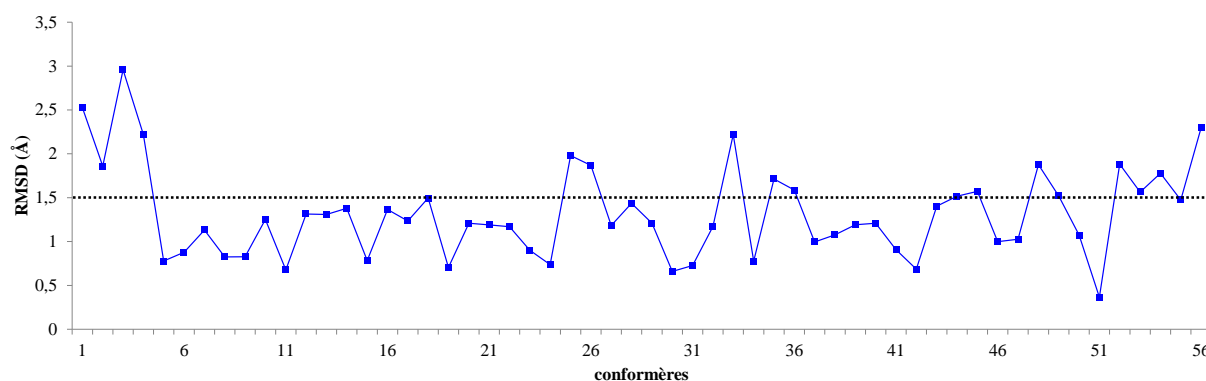


Figure III.2.5. Valeurs RMSD (Å) des géométries  $[\text{GB}_3 + 2\text{H}]^{2+}$  AMOEBA avec comme référence les géométries M06/6-31+G(d,p).

Pour résoudre ce problème, nous aurions pu optimiser directement toutes les structures générées lors de la dynamique moléculaire avec la méthode de calcul M06/6-31+G(d,p). Mais au vu de la taille des clusters étudiés, ce procédé n'était pas envisageable. Nous savons que les énergies AMOEBA peuvent être améliorées par des méthodes semi-empirique, PM6 ou PM6-DH2 par exemple, ce dernier s'étant avéré fiable lors de l'étude des acides aminés micro-hydratés au chapitre III partie 1. De ce fait, suivant le protocole défini précédemment, nous avons réalisé un calcul simple point d'énergie PM6-DH2 sur les géométries AMOEBA obtenues (voir les graphiques a) et b) de la Figure III.2.6). Les résultats n'indiquent cependant pas d'amélioration par rapport à AMOEBA. On peut penser que ce résultat décevant est lié à la mauvaise estimation des structures géométriques par AMOEBA, un simple calcul d'énergie sur ces structures n'étant alors pas suffisant. Nous avons donc ré-optimisé toutes les structures AMOEBA avec la méthode semi-empirique PM6. Comme le montre les graphiques c) et d) de la Figure III.2.6 et la Figure III.2.7, il y a un meilleur accord à la fois énergétique et géométrique entre PM6 et M06/6-31+G(d,p). En moyenne, le RMSD obtenu en AMOEBA et PM6 pour les structures  $[\text{GB}_3 + 2\text{H}]^{2+}$  est égale à 1,32 et 0,80 Å, respectivement. Seules quatre structures ont un RMSD supérieur à 1,5 Å avec la méthode de calcul PM6 (Figure III.2.7). On notera en particulier que les optimisations de géométrie PM6 et M06 permettent

des transferts de proton, ce qui est parfois observé ici, alors que les structures initiales AMOEBA ne pouvaient montrer ce transfert.

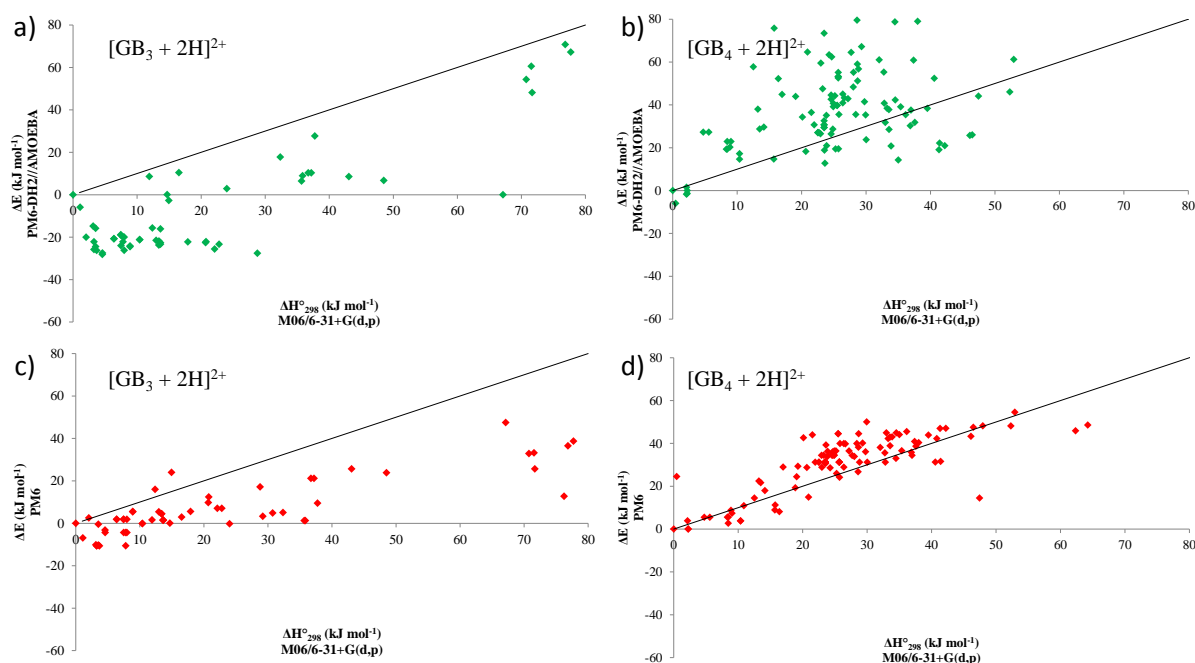


Figure III.2.6. Corrélation énergétique pour  $[\text{GB}_3 + 2\text{H}]^{2+}$  entre PM6-DH2//AMOEBA et M06/6-31+G(d,p) en haut de la figure et entre PM6 et M06/6-31+G(d,p) en bas de la figure (kJ mol<sup>-1</sup>).

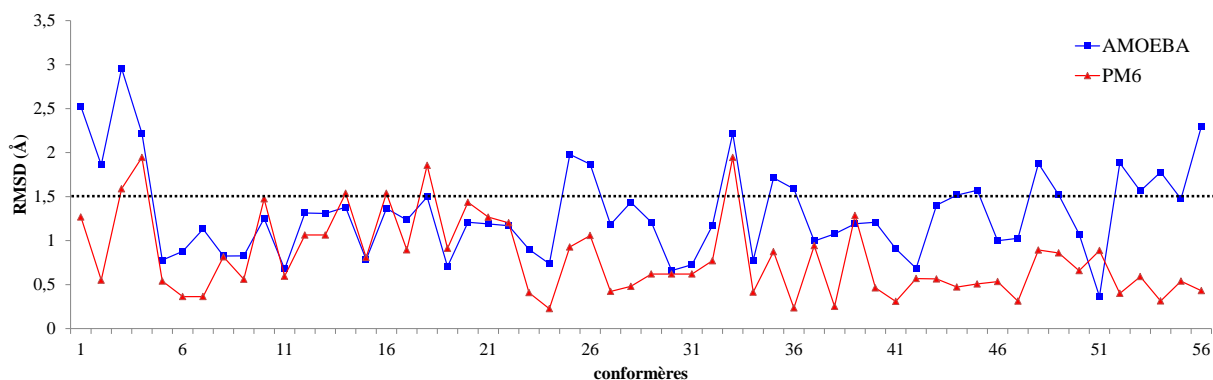


Figure III.2.7. Comparaison des RMSD (Å) avec les géométries  $[\text{GB}_3 + 2\text{H}]^{2+}$  AMOEBA et PM6 avec comme référence les géométries M06/6-31+G(d,p).

Ainsi, l'ajout d'une étape de calcul supplémentaire au niveau semi-empirique entre la phase d'optimisation des géométries AMOEBA et celle des optimisations avec M06/6-31+G(d,p) améliore sensiblement les résultats. On notera que lorsqu'une même structure est générée plusieurs fois en AMOEBA, celle-ci ne sera optimisée qu'une seule fois en PM6. Évidemment, s'il s'agit d'une nouvelle structure, sa géométrie sera optimisée en PM6. Outre



cette optimisation de géométrie en PM6, une autre différence dans notre stratégie de modélisation, par rapport à celle utilisée pour la micro-hydratation, intervient dans la suite. En effet, le choix du nombre de structures PM6 dont la géométrie doit être ré-optimisée en M06/6-31+G(d,p) pour chaque famille est un compromis entre critère énergétique et variété des structures des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ , l'objectif de ces travaux étant d'expliquer et de comprendre les tendances expérimentales observées. Ainsi, toutes les structures dans une gamme d'énergie donnée ne sont pas systématiquement recalculées, seules celles ayant une structure géométrique sensiblement différentes le sont. La stratégie d'exploration de surface des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  est représentée dans la Figure III.2.8. Cette stratégie a été appliquée pour  $[\text{GB}_3 + 2\text{H}]^{2+}$ ,  $[\text{GB}_4 + 2\text{H}]^{2+}$ ,  $[\text{GB}_5 + 2\text{H}]^{2+}$  et  $[\text{GB}_{13} + 2\text{H}]^{2+}$ . Pour ces deux dernières tailles de clusters, nous avons présenté uniquement les résultats principaux.

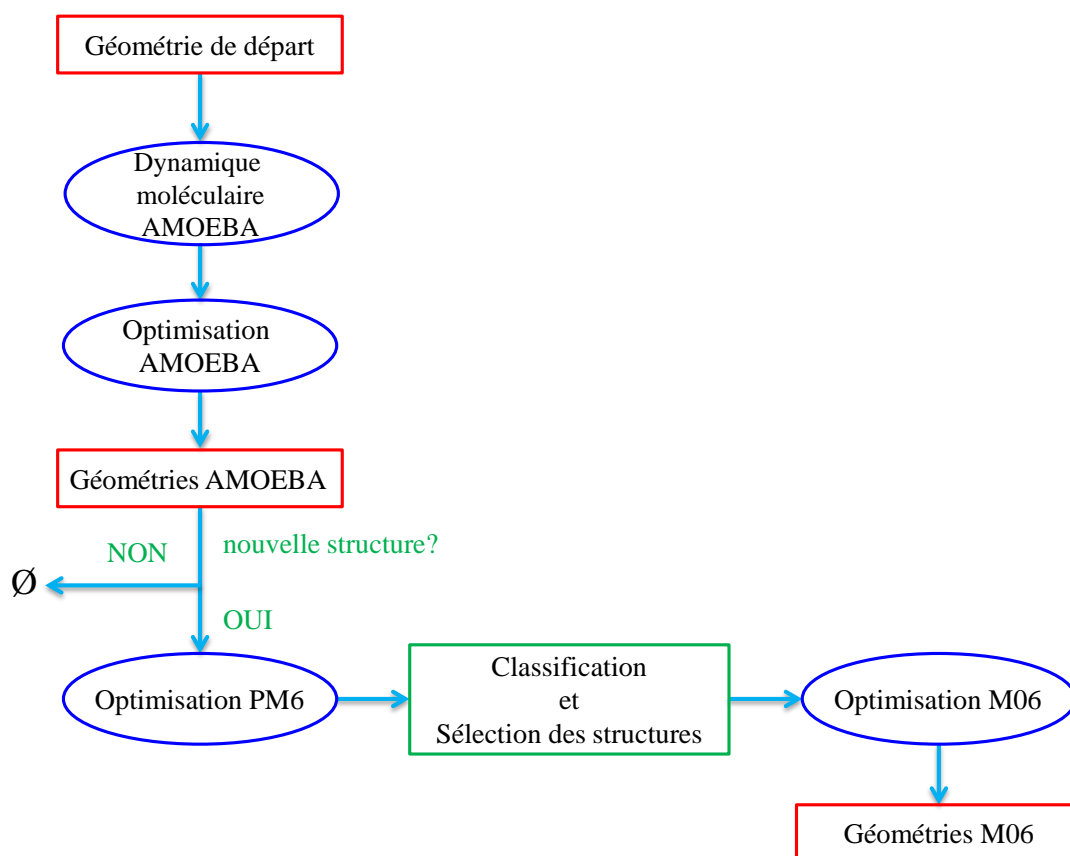


Figure III.2.8. Stratégie d'exploration de surface d'énergie potentielle.

### III.2.2.2. Base spécifique pour $[GB_{13} + 2H]^{2+}$

L'étude de ces clusters nécessite l'utilisation de bases incluant des fonctions de polarisation et des fonctions diffuses comme c'est le cas avec la base 6-31+G(d,p). Mais l'optimisation des structures du cluster  $[GB_{13} + 2H]^{2+}$  est très coûteuse en temps de calcul même avec cette base de Pople. Nous avons donc cherché à diminuer la taille de cette base de plusieurs façons lors du processus d'optimisation de géométrie de ce cluster uniquement. Les énergies obtenues ont ensuite été corrigées par calcul simple point M06/6-31+G(d,p). Deux bases ont été testées : M06/6-31G(d,p) qui n'a plus de fonction diffuse et M06/6-31G(d,p\*) pour laquelle, outre la suppression des fonctions diffuses, les fonctions de polarisation sur les atomes d'hydrogène des groupements méthyles ont été éliminées. Nous avons testé ces deux bases sur des critères géométriques et énergétiques sur quatre structures de  $[GB_3 + 2H]^{2+}$  (**1**, **2**, **3** et **4**) (Figure III.2.9).

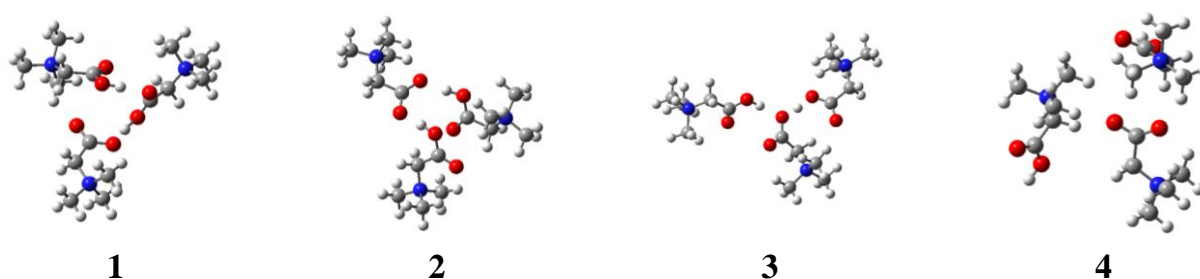


Figure III.2.9. Structures de  $[GB_3 + 2H]^{2+}$  utilisées pour les tests sur les bases atomiques (géométries M06/6-31+G(d,p)).

Les résultats obtenus sont reportés dans le Tableau III.2.2 et Figure III.2.10. De ces données, on constate que les géométries évoluent très peu, que l'on utilise M06/6-31G(d,p) ou M06/6-31G(d,p\*), comme l'atteste les valeurs RMSD qui valent en moyenne 0.144 et 0.149 Å respectivement. La même conclusion est obtenue d'un point de vue énergétique. En effet, pour les méthodes M06/6-31G(d,p), M06/6-31+G(d,p)/M06/6-31G(d,p), M06/6-31G(d,p\*) et M06/6-31+G(d,p)/M06/6-31G(d,p\*), la Déviation Moyenne Absolue (MAD) sur les énergies relatives vaut 1.2, 2.3, 0.9 et 2.4 kJ mol<sup>-1</sup>, respectivement. De plus, la MAD sur l'écart entre les énergies relatives a également été calculée et est égale à 1.07, 1.40, 1.40 et 1.43 kJ mol<sup>-1</sup>, respectivement. Donc pour les cinq niveaux de calcul utilisés, les tendances obtenues pour **1**, **2**, **3** et **4** sont les mêmes, avec **1** et **2** proches en énergies (écart énergétique inférieur à |1.5 kJ mol<sup>-1</sup>|).

Tableau III.2.2. Valeurs RMSD (Å) des structures **1**, **2**, **3** et **4**.

Structures	M06/6-31G(d,p)	M06/6-31G(d,p*)
<b>1</b>	0.315	0.342
<b>2</b>	0.134	0.129
<b>3</b>	0.043	0.039
<b>4</b>	0.085	0.084
Moyenne	0.144	0.149

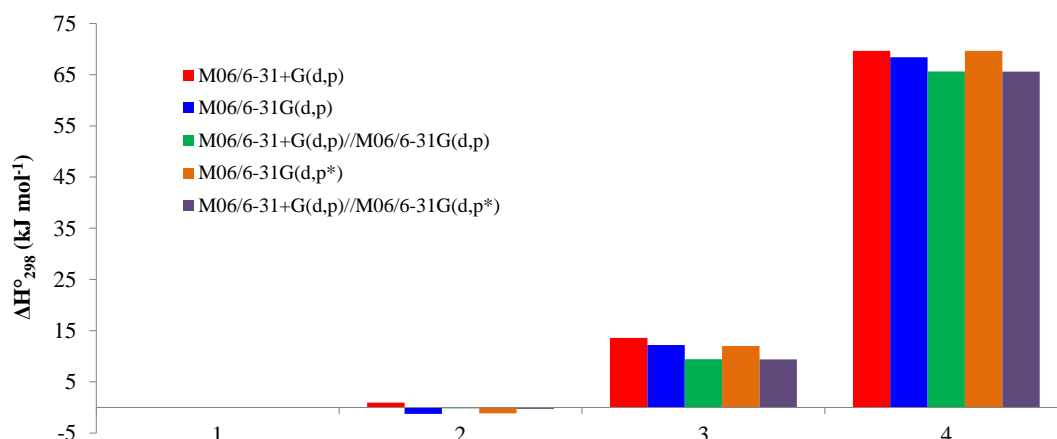


Figure III.2.10. Influences de la base sur les  $\Delta H^\circ_{298}$ . Les enthalpies relatives ( $\text{kJ mol}^{-1}$ ) sont calculées par rapport à **1**.

Puisque les résultats obtenus sont quasi-identiques pour M06/6-31G(d,p) et M06/6-31G(d,p\*), nous avons opté pour optimiser les structures sélectionnées du cluster  $[\text{GB}_{13} + 2\text{H}]^{2+}$  avec la base 6-31G(d,p\*) car elle est moins couteuse en temps de calcul. Même si les énergies M06/6-31+G(d,p) et M06/6-31G(d,p\*) sont similaires, un calcul simple point M06/6-31+G(d,p) sera systématiquement fait afin de pouvoir comparer les énergies de cohésion, d'élimination d'un neutre et de séparation de charge avec les clusters de plus petites tailles.

## III.2.3. Résultats et discussion

### III.2.3.1. Structures géométriques

Notre stratégie d'exploration de surface mise au point dans ce chapitre nous a permis d'identifier les différentes familles obtenues pour les clusters  $[\text{GB}_3 + 2\text{H}]^{2+}$ ,  $[\text{GB}_4 + 2\text{H}]^{2+}$  et  $[\text{GB}_5 + 2\text{H}]^{2+}$ . La structure la plus représentative a été sélectionnée pour chaque cluster  $[\text{GB}_n + 2\text{H}]^{2+}$  afin d'étudier la nature de leurs interactions intermoléculaires. L'ensemble des

structures identifiées est représenté dans la Figure III.2.11. À partir de  $[\text{GB}_3 + 2\text{H}]^{2+}$ , nous avons pu identifier plusieurs interactions intermoléculaires. Cinq motifs d'interaction (A), (B), (C), (D) et (E), répertoriés dans la Figure III.2.11, ont été trouvés pour ce système. Ces interactions ont été représentées dans la Figure III.2.12 pour le motif le plus stable nommé (A). Il y a au plus deux liaisons hydrogène de type  $\text{COOH} \cdots \text{OCO}$ , des interactions coulombiennes attractives ( $\text{COO}^- \cdots \text{N}^+$ ) et répulsives ( $\text{COO}^- \cdots \text{COO}^-$  et  $\text{N}^+ \cdots \text{N}^+$ ) qui induisent la disposition des GB dans le cluster.

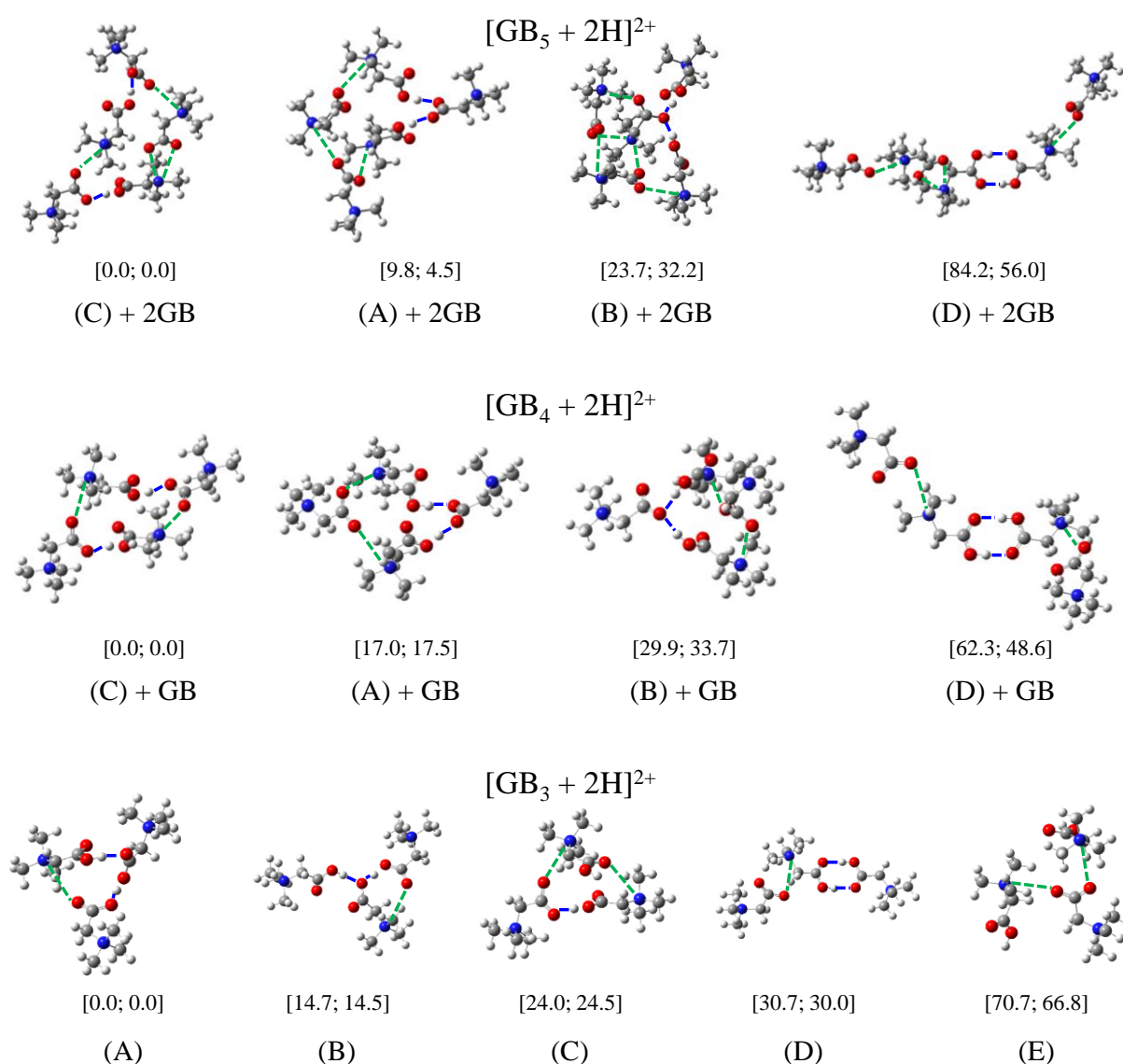


Figure III.2.11. Structures locales pour les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  avec  $n = 3$  à 5. Les valeurs entre crochet correspondent aux enthalpies ( $\Delta H^\circ_{298}$ ) et enthalpies libres relatives ( $\Delta G^\circ_{298}$ ) respectivement obtenus avec le niveau de théorie M06/6-31+G(d,p). Ces valeurs sont exprimées en  $\text{kJ mol}^{-1}$ .

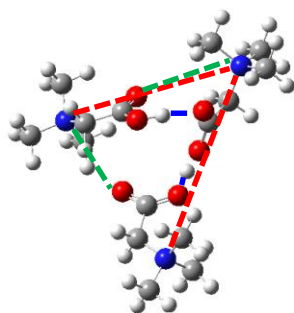


Figure III.2.12. Bilan des interactions intermoléculaires possibles au sein de ces clusters. Les lignes pointillées en rouge, vert et bleu représentent les interactions coulombiennes répulsives  $N^+ \cdots N^+$ , les interactions coulombiennes attractives  $COO^- \cdots N^+$  et les liaisons hydrogène  $COOH \cdots OCO$ .

(A) inclut deux liaisons hydrogène faisant intervenir les deux atomes d'oxygène d'une même GB et une ou plusieurs interactions coulombiennes du type  $COO^- \cdots N^+$  et  $HOCO^- \cdots N^+$ . Le motif (B) est similaire au motif (A) mais les deux liaisons hydrogène présentes impliquent le même atome d'oxygène de GB. Le motif (C) inclut la formation d'une seule liaison hydrogène dans le cas du trimère, et la glycine protonée libre va alors pouvoir former une liaison hydrogène avec une autre GB à partir de  $n = 4$ . Il y a donc la formation de deux liaisons hydrogène du type  $(GB \cdots H^+ GB)$  pour le motif (C) pour  $n > 3$ . Les deux  $[GB + H]^+$  formant deux liaisons hydrogène entre eux forment le motif (D). Le motif (E), impliquant uniquement des interactions coulombiennes attractives, est de plus en plus fréquent dans les autres familles lorsque le nombre de GB augmente. La diminution du nombre de liaisons hydrogène et le nombre de GB impliqué dans ces interactions permet d'expliquer l'ordre énergétique de ces différents motifs au sein de  $[GB_3 + 2H]^{2+}$ . L'augmentation de la taille des clusters  $[GB_n + 2H]^{2+}$ , c'est-à-dire pour  $n$  allant de 3 à 5, ne modifie pas les structures locales déjà identifiées dans le petit cluster  $[GB_3 + 2H]^{2+}$ . En effet, les motifs (A), (B), (C) et (D) se retrouvent dans les familles des clusters  $[GB_4 + 2H]^{2+}$  et  $[GB_5 + 2H]^{2+}$  car les GBs additionnels se lient ou s'insère au motif central. L'ordre énergétique ( $\Delta H^\circ_{298}$  et  $\Delta G^\circ_{298}$ ) pour les différents motifs restent le même d'une génération à une autre sauf pour le motif (C) qui devient le plus favorable à partir de  $n = 4$  ( $[GB_4 + 2H]^{2+}$ ,  $[GB_5 + 2H]^{2+}$  et  $[GB_{13} + 2H]^{2+}$ ). Ce n'est pas surprenant car l'ajout d'une GB pour ce motif entraîne la formation d'une liaison hydrogène stabilisante supplémentaire. L'écart énergétique est maintenu entre les motifs (A) et (B). Par contre une déstabilisation du motif (D) est observée. Pour  $[GB_3 + 2H]^{2+}$ ,  $[GB_4 + 2H]^{2+}$  et  $[GB_5 + 2H]^{2+}$ , le motif (D) est situé à 30.7, 45.3 et 74.4 kJ mol<sup>-1</sup> par rapport au motif (A).

Le cluster  $[\text{GB}_4 + 2\text{H}]^{2+}$  a été étudié par Feketeová et O'Hair au niveau de théorie PM3.[24] Trois structures possédant le motif (D) ont été considérées. Ils ont montré que ce motif est stable cinétiquement mais pas thermodynamiquement, et ont donc supposé que les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  pouvaient avoir un motif de ce type-là. Nous avons vu qu'en fait ce motif est l'un des moins stable, et donc il doit être peut fréquent par rapport à d'autres motifs que nous avons identifiés au cours de notre travail.

Concernant  $[\text{GB}_{13} + 2\text{H}]^{2+}$ , nous avons optimisé la géométrie de 7 clusters. Ceux-ci sont représentatifs des structures les plus stables identifiées présentant des caractéristiques différentes (motifs, distances et/ou localisation des deux  $[\text{GB} + \text{H}]^+$ ) (voir Tableau III.2.3) Nous avons représenté dans la Figure III.2.13 seulement trois de ces sept structures, G13\_1, G13\_2 et G13\_7, qui sont les plus intéressantes pour notre étude. On constate que parmi les structures les plus stables le motif (C) y est majoritairement inclut, une structure avec le motif (A) est plus haute en énergie, tandis que les motifs (B) et (D) n'ont pas été trouvé. L'ordre énergétique entre les motifs (C) et (A) est cohérent avec celui obtenu pour  $[\text{GB}_4 + 2\text{H}]^{2+}$  et  $[\text{GB}_5 + 2\text{H}]^{2+}$ . Toutefois, les énergies relatives augmentent significativement entre ces deux motifs ( $87.1 \text{ kJ mol}^{-1}$ ) ce qui est dû à l'attachement de huit GB additionnelles. Parmi les structures les plus stables, G13\_2 a pour particularité d'avoir les deux interactions  $\text{GB} \cdots \text{H}^+ \text{GB}$  très éloignées l'une de l'autre, tandis qu'elles sont proches dans le cas de G13\_1 comme l'attestent les distances  $\text{H}^+ \cdots \text{H}^+$  (voir Figure III.2.13). Les deux protons viennent neutraliser deux charges de type  $\text{COO}^-$  ; il n'est donc pas si étonnant à ce qu'ils soient à proximité dans la structure la plus stable. Afin de nous assurer que GB13\_1 est bien plus stable que GB13\_2, nous avons réalisé un calcul simple point en énergie M06/6-311++G(d,p) sur les géométries M06/6-31G(d,p\*) pour valider la qualité du niveau de théorie M06/6-31+G(d,p)//M06/6-31G(d,p\*). Les enthalpies et enthalpies libres relatives sont peu modifiées avec une différence de  $0.7 \text{ kJ mol}^{-1}$  seulement.

Tableau III.2.3. Enthalpie et enthalpie libre relatives associées aux structures de  $[\text{GB}_{13} + 2\text{H}]^{2+}$  ( $\text{kJ mol}^{-1}$ ).

Structures	Motif	PM6	M06/6-31+G(d,p)//M06/6-31G(d,p*)		M06/6-311++G(d,p)//M06/6-31G(d,p*)	
		$\Delta H^\circ_{298}$	$\Delta H^\circ_{298}$	$\Delta G^\circ_{298}$	$\Delta H^\circ_{298}$	$\Delta G^\circ_{298}$
GB13_1	(C)	0.0	0.0	0.0	0.0	0.0
GB13_2	(C)	6.0	13.8	13.0	14.5	13.7
GB13_3	(C)	6.7	52.0	26.2	/	/
GB13_4	(C)	11.3	23.9	23.6	/	/
GB13_5	(C)	11.6	20.4	39.2	/	/
GB13_6	(C)	21.0	59.6	60.2	/	/
GB13_7	(A)	42.5	80.8	87.1	/	/

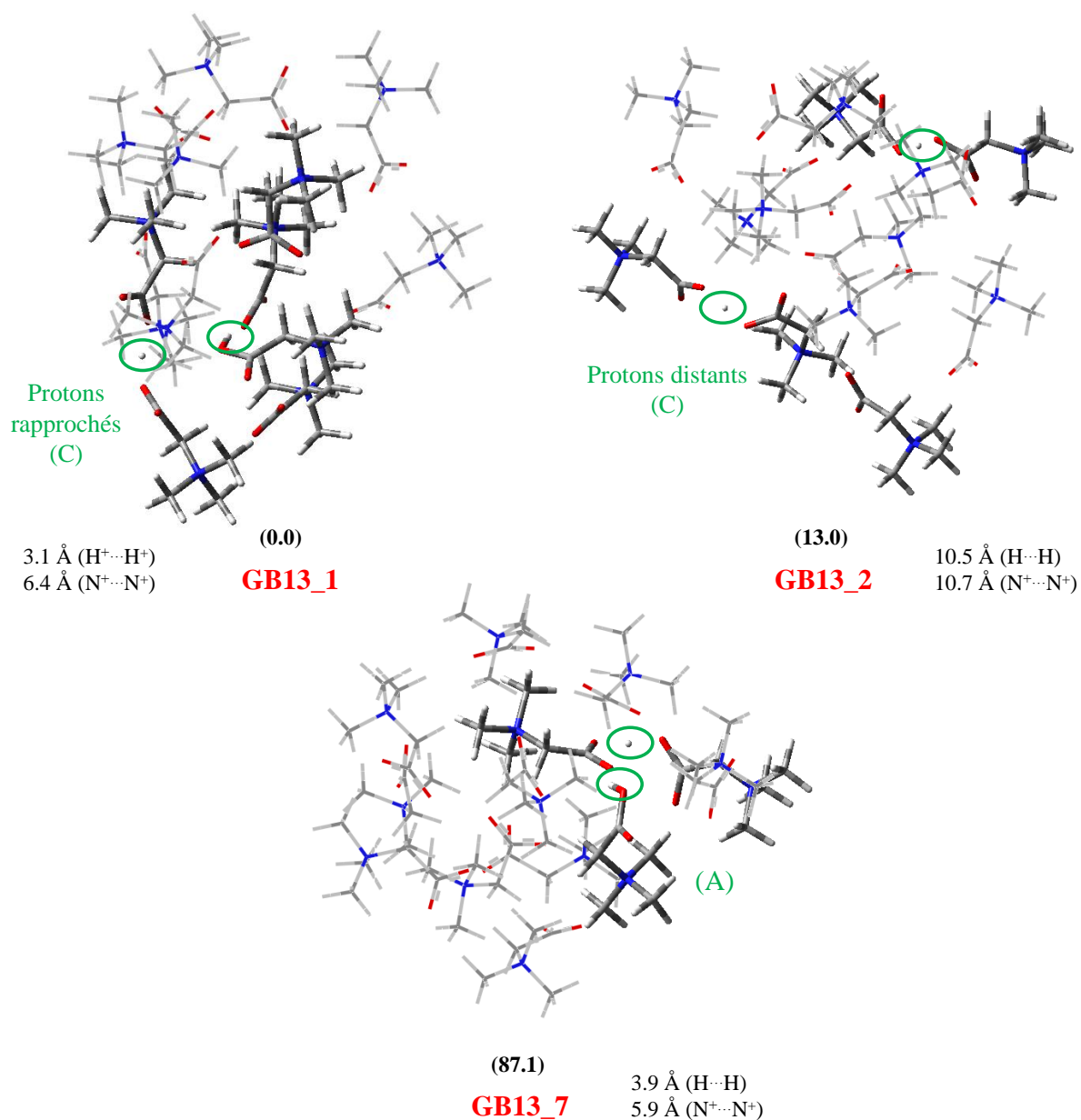


Figure III.2.13. Structures de  $[\text{GB}_{13} + 2\text{H}]^{2+}$ . Les valeurs entre parenthèse correspondent aux enthalpies libres obtenues au niveau de théorie M06/6-31+G(d,p)//M06/6-31G(d,p\*) ( $\text{kJ mol}^{-1}$ ).

Lorsque l'on visualise les structures GB13\_1 et GB13\_2, on constate que ces deux clusters ont une organisation structurale de type feuillet. Les découpages en feuillets sont représentés dans la Figure III.2.14 ; ces feuillets sont à priori constitués d'une succession d'interactions coulombiennes de type  $\text{COO}^-\cdots\text{N}^+$  principalement et sont reliés les uns aux autres par ce même type d'interaction.

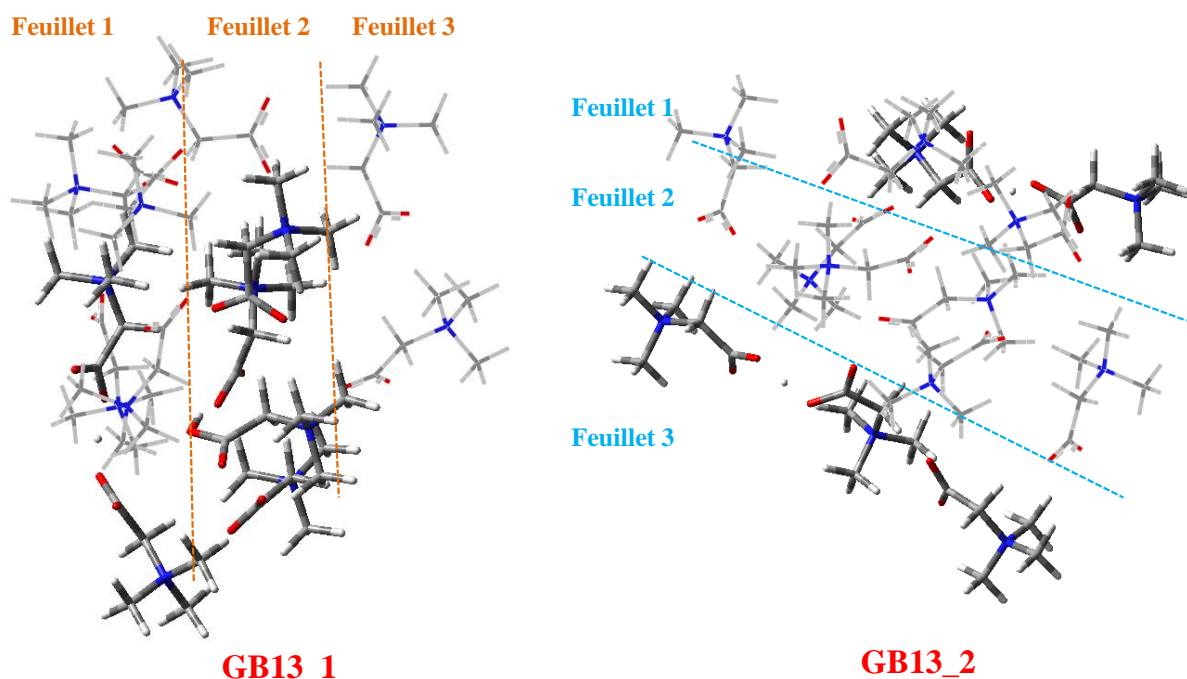


Figure III.2.14. Arrangement en feuillet des structures GB13\_1 et GB13\_2.

De plus, nous avons cherché à identifier la plus petite unité  $[\text{GB}_n + 2\text{H}]^{2+}$  permettant de décrire au mieux GB13\_1. Parmi les structures locales possibles, nous avons extrait le motif (C) en considérant un tétramère, un pentamère et un hexamère (voir Figure III.2.15). Ces structures seules ont été optimisées au niveau M06/6-31G(d,p\*) et comparées à leurs géométries dans le cluster, en cherchant celui dont la disposition change légèrement par rapport à celui dans le clusters. Au vu des structures représentées dans la Figure III.2.15 et des valeurs RMSD obtenues, 4.537, 3.235 et 1.748 Å respectivement, nous pouvons déduire que GB13\_1 intègre un motif  $[\text{GB}_6 + 2\text{H}]^{2+}$ . Cette structure sera utilisée dans les parties III.2.3.2.2 et III.2.3.3 pour l'analyse des énergies d'interactions.



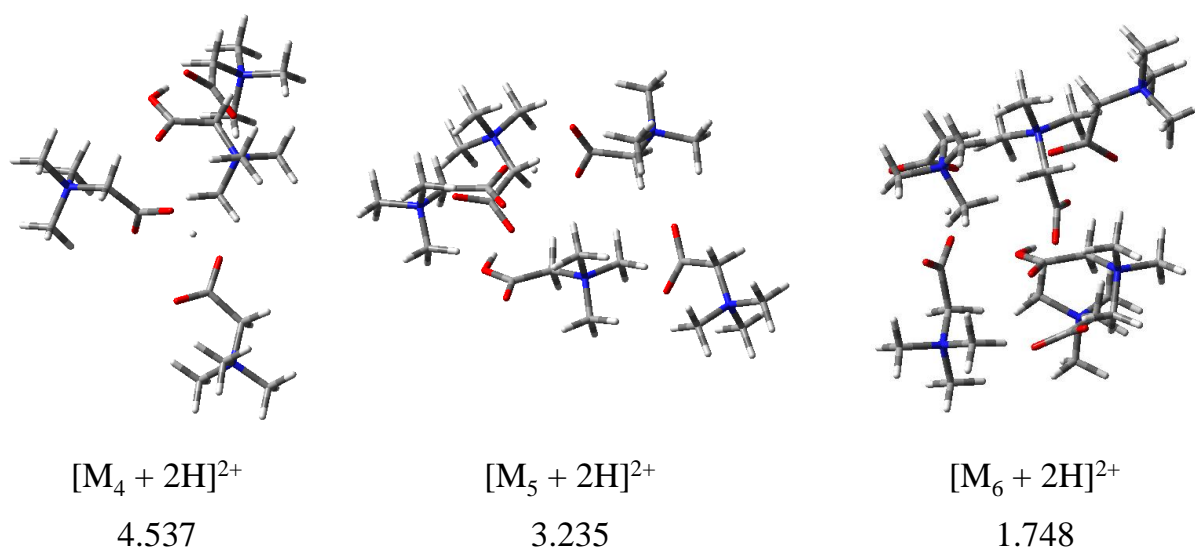


Figure III.2.15. Identification du plus petit motif dans GB13\_1. Géométries optimisées avec M06/6-31+G(d,p). Valeurs RMSD en Å.

Finalement,  $[GB_{21} + 2H]^{2+}$  est un cluster de très grande taille ne permettant pas pour un temps de calcul raisonnable d'explorer exhaustivement sa surface d'énergie potentielle avec notre méthodologie. Nous avons donc utilisé GB13\_1 et GB13\_2 pour construire manuellement  $[GB_{21} + 2H]^{2+}$  en ajoutant des GB de façon à faire varier la localisation des protons au sein des clusters (centres ou extrémités du cluster). Les structures obtenues ont été optimisées avec PM6. Au sein de GB21\_1, les deux protons sont proches et ils sont situés au centre du cluster ; tandis qu'avec GB21\_2, les protons sont éloignés et l'un d'eux est situé au centre du cluster, et l'autre est à l'extrémité. Les deux structures les plus stables obtenues, GB21\_1 et GB21\_2, ont des enthalpies relatives PM6 égales à 0.0 et 46.1 kJ mol<sup>-1</sup> respectivement (voir Figure III.2.16). Cela est cohérent avec la tendance observée précédemment dans laquelle l'addition de plusieurs GB entraîne l'augmentation de l'écart énergétique entre les différents conformères.

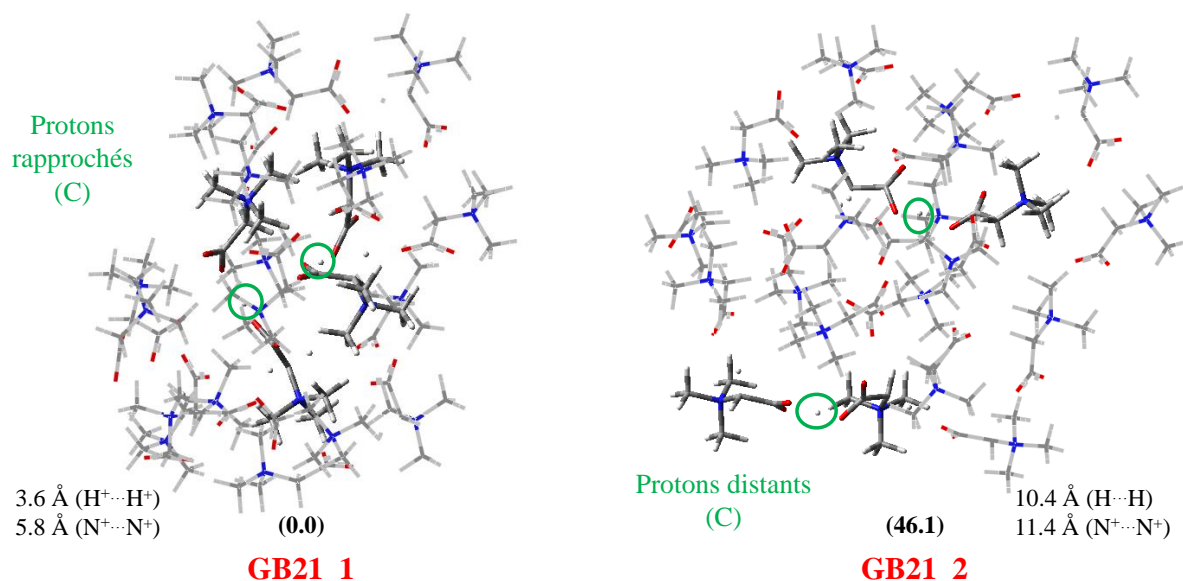


Figure III.2.16. Les deux structures de  $[\text{GB}_{21} + 2\text{H}]^{2+}$  les plus stables identifiées. Géométries optimisées en PM6. Les valeurs entre parenthèse correspondent aux énergies obtenues au niveau de théorie PM6 ( $\text{kJ mol}^{-1}$ ).

### III.2.3.2. Nature des interactions non-covalentes intermoléculaires

#### III.2.3.2.1. Interactions non-covalentes dans $[\text{GB}_n + 2\text{H}]^{2+}$

Plusieurs méthodes ont été employées pour l'étude de diverses interactions non-covalentes présentes dans nos systèmes. En utilisant la méthode NCI, nous avons voulu vérifier si seules les interactions énumérées dans la partie III.2.3.1 sont présentes, ou bien si d'autres interactions viennent s'ajouter à cette liste. Nous avons utilisé les densités SCF lors des calculs NCI.

Notre investigation commence par l'analyse des différents types d'interactions dans différents dimères de glycine bêtaïne, neutre ( $[\text{GB}_2]$ ), mono- ( $[\text{GB}_2 + \text{H}]^+_1$  et  $[\text{GB}_2 + \text{H}]^+_2$ ) et diprotoné ( $[\text{GB}_2 + 2\text{H}]^{2+}$ ). Pour cela, ces dimères ont été construits avec l'intuition chimique de façon à reproduire les différents types d'interactions déjà observées dans  $[\text{GB}_n + 2\text{H}]^{2+}$  et par conséquent identifier de nouvelles interactions. Pour  $[\text{GB}_2 + 2\text{H}]^{2+}$ , on distingue deux liaisons hydrogène ( $\text{sign}(\lambda_2)\rho = -0.034$ ) entre les deux acides carboxyliques ; à cette interaction s'ajoute une faible interaction de van der Waals entre ces deux hydrogène acides (voir Figure III.2.17). Ces résultats sont en accord avec les interactions non-covalentes

présentes dans le dimère de l'acide formique.[28] Les interactions entre le groupe COOH et le groupement ammonium sont nombreuses ; il existe au sein d'une même GB (voir Figure III.2.17) :

- une interaction contrainte du type charge...dipôle ( $\text{HOCO} \cdots \text{N}^+$ ) (vert)
- des répulsions stériques entre CO et les groupements méthyles (rouge)
- des faibles liaisons hydrogène  $\text{HOCO} \cdots \text{HC}_{\text{methyl}}$  ( $\text{sign}(\lambda_2)\rho = -0.012$ ) (bleu).

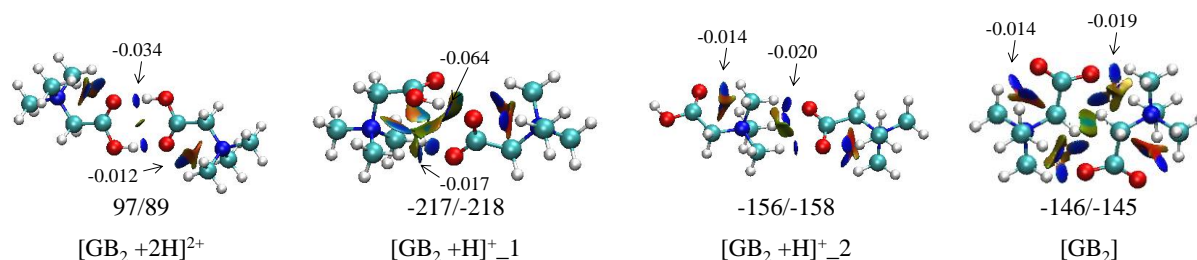


Figure III.2.17. Analyse NCI au sein de dimère neutre, mono et di-protoné de glycine bétaine (géométries M06/6-31+G(d,p)). L'échelle de couleur est  $-0.015 < \text{sign}(\lambda_2)\rho < 0.015$  au. Les valeurs associées en bas des figures sont les énergies de cohésion (voir partie III.2.3.2.2) calculées au niveau M06/6-31+G(d,p) et B3LYP-D3/TZP respectivement en  $\text{kJ mol}^{-1}$ . Les autres valeurs représentent les valeurs de  $\text{sign}(\lambda_2)\rho$  définies dans le chapitre II partie 9.2.

Des interactions similaires ont été trouvées dans presque toutes les GB et  $[\text{GB} + \text{H}]^+$  (pour exception voir le modèle (a) représenté dans la Figure III.2.18). Les interactions présentes dans  $[\text{GB}_2 + \text{H}]^+_1$  sont plus difficiles à identifier excepté la liaison hydrogène  $\text{COOH} \cdots \text{OCO}$  ( $\text{sign}(\lambda_2)\rho = -0.064$ ) et les deux liaisons hydrogène intermoléculaires  $\text{OCO} \cdots \text{HC}_{\text{methyl}}$  ( $\text{sign}(\lambda_2)\rho = -0.017$ ). Pour comprendre la provenance de ces interactions, nous avons substitué certains groupements par un ou plusieurs atomes d'hydrogène. Les modèles utilisés sont reportés dans la Figure III.2.18. Le modèle (a) permet de mettre en évidence l'existence de faibles interactions de type (1) liaison hydrogène intramoléculaire:  $\text{HOCO} \cdots \text{HC}_{\text{methyl}}$  et  $\text{HO} \cdots \text{HC}_{\text{methyl}}$  (bleu), (2) charge-dipôle entre  $\text{N}^+$  et COOH au sein de  $[\text{GB} + \text{H}]^+$  (vert) et (3) répulsion stérique intramoléculaire entre COOH et les groupements méthyles (orange) qui ont déjà été identifiées. Selon les modèles (b) et (c), une faible interaction de van der Waals et une répulsion stérique existent en plus entre COOH et  $\text{COO}^-$ , noté  $\text{COOH} \cdots \text{COO}^-$ .

$[\text{GB}_2 + \text{H}]^+_2$  est stabilisé par une interaction charge-charge entre  $\text{COO}^- \cdots \text{N}^+$  (vert) mais également par trois liaisons hydrogène intermoléculaires  $\text{OCO} \cdots \text{HC}_{\text{methyl}}$  ( $\text{sign}(\lambda_2)\rho = -0.020$ , bleu). Évidemment, il y a de la répulsion stérique entre  $\text{COO}^-$  (GB) et les groupements méthyles ( $[\text{GB} + \text{H}]^+$ ) (orange).

Entre deux GB ([GB<sub>2</sub>]), il existe en plus des interactions déjà citées (c'est-à-dire les liaisons hydrogène  $\text{OCO}^{\cdots}\text{HC}_{\text{methyl}}$  intra ( $\text{sign}(\lambda_2)\rho = -0.014$ , bleu) et intermoléculaires ( $\text{sign}(\lambda_2)\rho = -0.019$ , bleu), les interactions  $\text{COO}^{\cdots}\text{N}^+$  (vert) et les répulsions stériques entre CO et les groupements méthyles (rouge et orange)) une interaction de type van der Waals entre les CH<sub>2</sub> (vert) et deux liaisons hydrogène de type  $\text{COO}^{\cdots}\text{HC}_{\text{CH}_2}$ .

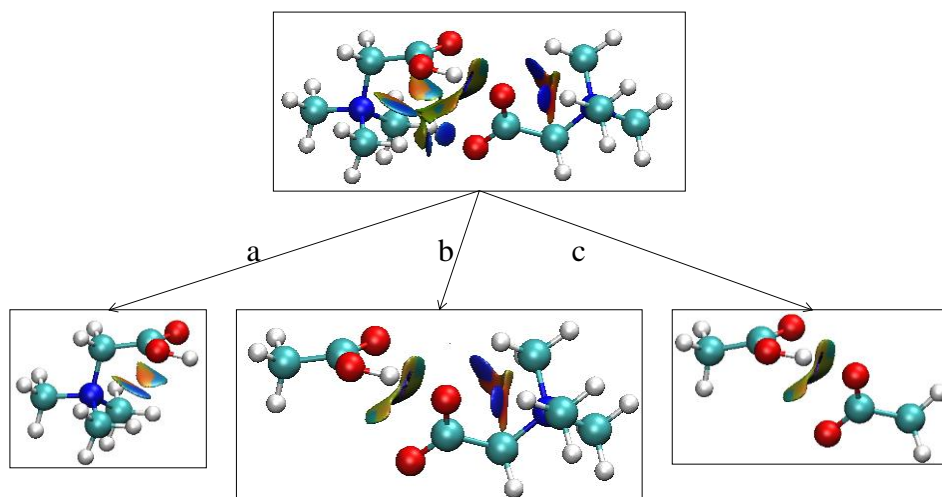


Figure III.2.18. Analyse NCI des modèles pour  $[\text{GB}_2 + \text{H}]^+_1$ . L'échelle de couleur est  $-0.015 < \text{sign}(\lambda_2)\rho < 0.015$  au.

Nous avons réalisé l'analyse NCI sur les motifs des clusters  $[\text{GB}_3 + 2\text{H}]^{2+}$  : (A), (C), (D) et (E) (Figure III.2.19). Nous avons vu que les interactions non-covalentes observées dans les modèles des dimères de glycine bêtaïne sont présentes dans ces systèmes. En général, l'addition ou l'insertion d'une GB entraîne des interactions de van der Waals supplémentaires stabilisantes.

Nous pouvons compléter notre analyse des interactions non-covalentes effectuée dans la partie III.2.3.1 pour les motifs du cluster  $[\text{GB}_3 + 2\text{H}]^{2+}$ . En plus des liaisons hydrogène autres que  $\text{COOH}^{\cdots}\text{OCO}$  et  $\text{COOH}^{\cdots}\text{OCOH}$ , les motifs (A) et (C) ont des interactions de van der Waals de type  $\text{COOH}^{\cdots}\text{COOH}$  et (ou)  $\text{COOH}^{\cdots}\text{COO}^-$ . Il existe en plus dans le motif (E) des interactions de van der Waals entre les groupements CH<sub>3</sub> de GB différentes.

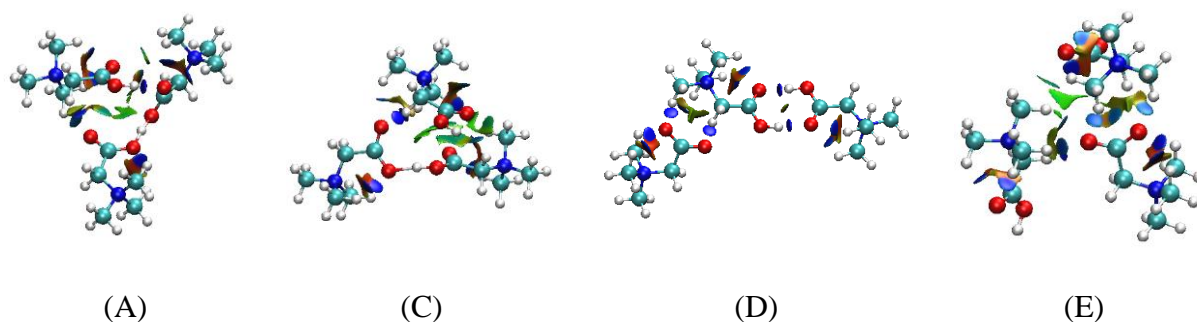


Figure III.2.19. Analyse NCI au sein des différents motifs de  $[\text{GB}_3 + 2\text{H}]^{2+}$  (géométries M06/6-31+G(d,p)). L'échelle de couleur est  $-0.015 < \text{sign}(\lambda_2)\rho < 0.015$  au.

Comme le montre la Figure III.2.20, les clusters  $[\text{GB}_{13} + 2\text{H}]^{2+}$  sont principalement stabilisés par les interactions de type (1) van der Waals entre  $\text{CH}\cdots\text{CH}$  (vert), (2) liaisons hydrogène  $\text{OCO}\cdots\text{HC}_{\text{CH}_2/\text{methyl}}$  en addition des deux liaisons hydrogène  $\text{COOH}\cdots\text{OCO}$  (bleu) et (3) charge-charge ( $\text{COO}^-\cdots\text{N}^+$ ) (vert). Nous pouvons généraliser ses observations aux clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  de plus grandes tailles.

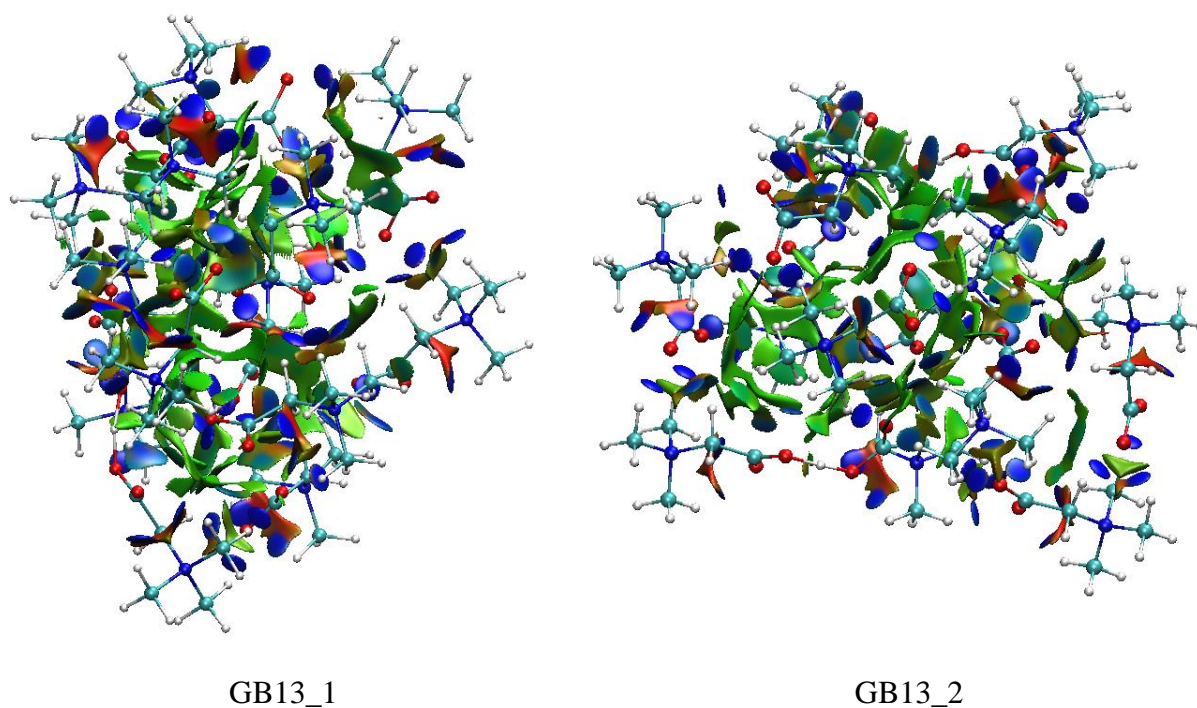


Figure III.2.20. Analyse NCI au sein de GB13\_1 et GB13\_2 (géométries M06/6-31G(d,p\*)). L'échelle de couleur est  $-0.015 < \text{sign}(\lambda_2)\rho < 0.015$  au.

### III.2.3.2.2. Cohésion au sein des clusters $[\text{GB}_n + 2\text{H}]^{2+}$

Maintenant que nous avons identifié les structures de certains clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  et la nature de leurs interactions non-covalentes, nous allons nous intéresser aux énergies d'interaction intermoléculaires afin de comprendre le comportement de ces clusters. Cette section a donc pour objectif de comprendre en partie certaines tendances expérimentales mentionnées dans la partie III.2.1.2.2. Nous avons choisi d'utiliser les énergies électroniques car la partition de l'énergie est effectuée en énergie électronique. De plus, les ordres relatifs des différents motifs ne sont pas modifiés en considérant  $\Delta E$  et  $\Delta H^\circ_{298}$ . Par exemple pour  $[\text{GB}_3 + 2\text{H}]^{2+}$ ,  $\Delta E = 0.0, 12.1, 18.5, 23.9$  et  $62.6 \text{ kJ mol}^{-1}$  pour les motifs, (A), (B), (C), (D) et (E), respectivement.

Nous avons calculé l'énergie d'interaction intermoléculaire totale ( $E_{int}$ ) au sein des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  associée à l'équation suivante :



La formation du cluster est favorable si  $E_{int} < 0$ , i.e que la réaction est exothermique. Il y a deux façons de calculer cette donnée. La première, noté *S1*, consiste à conserver la géométrie des différents fragments au sein du cluster, on parlera alors d'énergie de cohésion ( $E_{cohésion}$ ). La seconde, noté *S2*, consiste à prendre en compte la géométrie optimisée des fragments isolés, on parlera alors d'énergie de liaison ( $E_{liaison}$ ). Les énergies d'interactions intermoléculaires devraient être plus faibles dans le cadre *S2* que *S1* à cause de la pénalité énergétique due à la déformation des GB pour avoir la géométrie dans le cluster.

Dans un premier temps, dans le cadre *S1*, nous avons calculé cette énergie d'interaction ( $E_{cohésion}$ ) au sein des différents motifs de  $[\text{GB}_3 + 2\text{H}]^{2+}$ . Pour (A), (B), (C), (D) et (E) elle est égale à -222, -136, -126, -109, -67  $\text{kJ mol}^{-1}$  respectivement au niveau M06/6-31+G(d,p) (voir Tableau III.2.4). Ces valeurs ont été recalculées au niveau M06/6-311++G(d,p) sur les géométries M06/6-31+G(d,p), et elles diffèrent des valeurs M06/6-31+G(d,p) de 0 à 2  $\text{kJ mol}^{-1}$  seulement. Par un raisonnement similaire, le calcul des énergies de liaison totales est également suffisant avec le niveau de calcul M06/6-31+G(d,p). Donc dans les parties III.2.3.3.1 et III.2.3.3.2, nous nous contenterons du niveau de théorie M06/6-31+G(d,p). En M06/6-31+G(d,p) et B3LYP-D3/TZP, nous constatons que la cohésion entre les GB et  $[\text{GB} + \text{H}]^+$  est plus forte suivant cette ordre (A) > (B) > (C) > (D) > (E). De plus, cette énergie de cohésion totale varie de la même façon que les énergies relatives de ces motifs pour  $[\text{GB}_3 +$

$2\text{H}]^{2+}$ . Comme attendue, les énergies de liaisons sont plus faibles que les énergies de cohésion mais les tendances obtenues entre les approximations *S1* et *S2* restent similaires.

Tableau III.2.4. Énergies de cohésion et énergies de liaison totales associées à  $[\text{GB}_3 + 2\text{H}]^{2+}$  ( $\text{kJ mol}^{-1}$ ).\*

$[\text{GB}_3 + 2\text{H}]^{2+}$	M06/6-31+G(d,p)	M06/6-311++G(d,p)/M06/6-31+G(d,p)	B3LYP-D3/TZP
(A)	-222 (-117)	-222 (-114)	-222
(B)	-136 (-105)	-138 (-105)	-150
(C)	-126 (-98)	-128 (-98)	-136
(D)	-109 (-93)	-110 (-92)	-119
(E)	-67 (-54)	-68 (-53)	-75

\*Les valeurs X (Y) correspondent à  $E_{\text{cohésion}}$  ( $E_{\text{liaison}}$ ).

Précédemment, nous avons mis l'accent sur le fait que les motifs étaient conservés lors de l'augmentation de la taille du cluster  $[\text{GB}_n + 2\text{H}]^{2+}$ . Pour rappel, lorsque le cluster augmente en taille, (C) devient le motif le plus stable à partir de  $n = 4$ . Le Tableau III.2.5 regroupe les énergies de cohésions et de liaisons totales des clusters présentant ce motif, l'exception étant les petits clusters  $[\text{GB}_2 + 2\text{H}]^{2+}$  et  $[\text{GB}_3 + 2\text{H}]^{2+}$  (motif (D) et (A), respectivement). Nous n'avons pas considéré  $[\text{GB}_{21} + 2\text{H}]^{2+}$  dans cette partie et les suivantes car l'analyse des interactions non-covalentes et des processus de fragmentation se complexifie davantage avec la taille du cluster. Ces données permettent de confirmer que le niveau de théorie M06/6-31+G(d,p) est suffisant pour les calculs de  $E_{\text{cohésion}}$  et  $E_{\text{liaison}}$  puisque des écarts allant de 0 à 4  $\text{kJ mol}^{-1}$  par rapport à M06/6-311++G(d,p) ont été observés pour  $n = 2$  à 6, et de 5 à 10  $\text{kJ mol}^{-1}$  pour  $n = 13$ . De plus, les écarts énergétiques obtenus avec ces deux bases ne changent pas les énergies relatives des différents motifs.

$E_{\text{int}}([\text{GB}_2 + 2\text{H}]^{2+})$  est positive, ce qui implique que ce cluster ne peut pas exister. Cela est cohérent avec le fait que ce cluster n'a pas été détecté expérimentalement. Il n'est pas étonnant de constater que la cohésion au sein des clusters augmente avec leur taille. En divisant cette énergie par le nombre de GB neutre dans le cluster, cela nous permet d'estimer l'énergie moyenne apportée par l'addition d'une GB. Il est plaisant de constater que cette énergie de cohésion (liaison) additionnelle vaut  $\sim -180$  ( $-160$ )  $\text{kJ mol}^{-1}$  avec M06/6-31+G(d,p) et B3LYP-D3/TZP à partir de  $n = 4$ .

D'après nos observations, l'ajout d'une GB neutre apporte approximativement une interaction coulombienne attractive  $\text{COO}^- \cdots \text{N}^+$  et une qui est répulsive  $\text{N}^+ \cdots \text{N}^+$ . Nous avons estimé grâce à un modèle représenté à la Figure III.2.21 l'énergie due (i) à la formation d'une liaison hydrogène ( $E_{\text{H}}$ ), (ii) à la répulsion entre les deux ammoniums ( $E_{\text{NN}}$ ) et (iii) à l'attraction entre



O<sup>-</sup> et N<sup>+</sup> (E<sub>NO</sub>). Pour cela nous avons remplacé certains groupements par des atomes d'hydrogène. Ainsi, ajouter une GB neutre supplémentaire apporte approximativement une énergie de E<sub>NN</sub> + E<sub>NO</sub> = -176 kJ mol<sup>-1</sup>, ce qui est en accord avec le gain énergétique pour la cohésion provenant de le Tableau III.2.5. Mais les énergies de cohésion et de liaison totales ne permettent pas d'expliquer pourquoi le plus petit cluster observé est le [GB<sub>13</sub> + 2H]<sup>2+</sup>.

Tableau III.2.5. Énergies de cohésion et énergies de liaisons totales associées à [GB<sub>n</sub> + 2H]<sup>2+</sup> (kJ mol<sup>-1</sup>).\*

[GB <sub>n</sub> + 2H] <sup>2+</sup>	M06/6-31+G(d,p)		M06/6-311++G(d,p)/M06/6-31+G(d,p)		B3LYP-D3/TZP	
	<i>E<sub>int</sub></i>	$\frac{E_{int}}{n-2}$	<i>E<sub>int</sub></i>	$\frac{E_{int}}{n-2}$	<i>E<sub>int</sub></i>	$\frac{E_{int}}{n-2}$
[GB <sub>2</sub> + 2H] <sup>2+</sup>	97 (104)	/	96 (104)	/	89	/
[GB <sub>3</sub> + 2H] <sup>2+</sup>	-222 (-117)	-222 (-117)	-222 (-114)	-222 (-114)	-222	-222
[GB <sub>4</sub> + 2H] <sup>2+</sup>	-352 (-316)	-176 (-158)	-355 (-317)	-178 (-159)	-366	-183
[GB <sub>5</sub> + 2H] <sup>2+</sup>	-522 (-482)	-174 (-161)	-526 (-481)	-175 (-160)	-542	-181
[GB <sub>6</sub> + 2H] <sup>2+</sup>	-702 (-627)	-176 (-157)	-706 (-625)	-176 (-156)	-719	-180
[GB <sub>13</sub> + 2H] <sup>2+</sup>						
G13_1 <sup>#</sup>	-1973 (-1726)	-179 (-157)	-1978 (-1717)	-180 (-156)	-2014	-183
G13_2 <sup>#</sup>	-1993 (-1706)	-181 (-155)	-2001 (-1696)	-182 (-154)	-2039	-185

\* Les valeurs X (Y) correspondent à E<sub>cohésion</sub> (E<sub>liaison</sub>). <sup>#</sup> Géométries M06/6-31G(d,p\*).

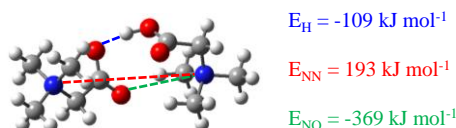


Figure III.2.21. Types d'interaction intermoléculaire et énergie associée. Énergies M06/6-31+G(d,p) en kJ mol<sup>-1</sup>.

Nous avons décomposé l'énergie de cohésion totale au sein de ces clusters pour identifier les contributions majoritaires avec le niveau de théorie B3LYP-D3/TZP. Pour rappel, celle-ci a pour expression (voir chapitre II):

$$E_{int} = \Delta E_{elect} + \Delta E_{Pauli} + \Delta E_{orb} + \Delta E_{disp} \quad (\text{III.2.13})$$

Le terme de dispersion ne sera pas discuté ici mais il contribue à la stabilisation de ces clusters. Il vaut -12, -43, -72, -107, -197, -643 et -602 kJ mol<sup>-1</sup> pour [GB<sub>2</sub> + 2H]<sup>2+</sup>, [GB<sub>3</sub> + 2H]<sup>2+</sup>, [GB<sub>4</sub> + 2H]<sup>2+</sup>, [GB<sub>5</sub> + 2H]<sup>2+</sup>, [GB<sub>6</sub> + 2H]<sup>2+</sup>, GB13\_1 et GB13\_2 respectivement, ce qui est en accord avec l'augmentation des interactions de van der Waals observée lors de l'analyse NCI dans la partie III.2.3.2.1.



Pour identifier la contribution énergétique dominante entre l'électrostatique et la répulsion de Pauli au sein des modèles  $[\text{GB}_2 + 2\text{H}]^{2+}$ ,  $[\text{GB}_2 + \text{H}]^+_1$ ,  $[\text{GB}_2 + \text{H}]^+_2$  et  $[\text{GB}_2]$ , nous avons quantifier la répulsion stérique avec B3LYP-D3/TZP en décomposant leur énergie d'interaction. La répulsion stérique est la somme des contributions électrostatiques et de la répulsion de Pauli et elle vaut 179, -77, -89 et -40  $\text{kJ mol}^{-1}$ , respectivement. Les valeurs négatives impliquent que les interactions électrostatiques sont suffisamment importantes pour contrer la répulsion de Pauli.

La Figure III.2.22 présente les interactions de Pauli, les interactions électrostatiques, les interactions orbitales et la répulsion stérique apportés par une GB pour chaque cluster le plus stable de chaque  $[\text{GB}_n + 2\text{H}]^{2+}$ , c'est à dire  $\frac{\Delta E_{\text{Pauli}}}{n-2}$ ,  $\frac{\Delta E_{\text{elect}}}{n-2}$ ,  $\frac{\Delta E_{\text{orb}}}{n-2}$  et  $\frac{\Delta E_{\text{ster}}}{n-2}$  respectivement. Ces différentes contributions se stabilisent à partir de  $n = 4$  car le motif le plus stable change à  $n = 4$  ((A) pour  $n = 3$ , et (C) à partir de  $n = 4$ ). Les tendances observées montrent que l'addition d'une GB apporte les mêmes contributions énergétiques au sein des clusters, ce qui justifie la valeur constante de  $\sim -180 \text{ kJ mol}^{-1}$  obtenue pour la série des  $[\text{GB}_n + 2\text{H}]^{2+}$ . Excepté pour  $[\text{GB}_3 + 2\text{H}]^{2+}$ , la contribution électrostatique est la plus favorable. Cela est confortant dû fait que ces clusters possèdent de nombreuses liaisons hydrogène, d'interactions type charge-charge et de van der Waals. La répulsion stérique est négative pour  $n > 3$  ce qui implique que le terme électrostatique prédomine sur la répulsion de Pauli. Pour  $[\text{GB}_3 + 2\text{H}]^{2+}$ , clusters non observés expérimentalement, la répulsion stérique est positive, ce qui implique que les forces électrostatiques ne sont pas suffisamment élevées pour permettre à elles seules la cohésion du cluster.

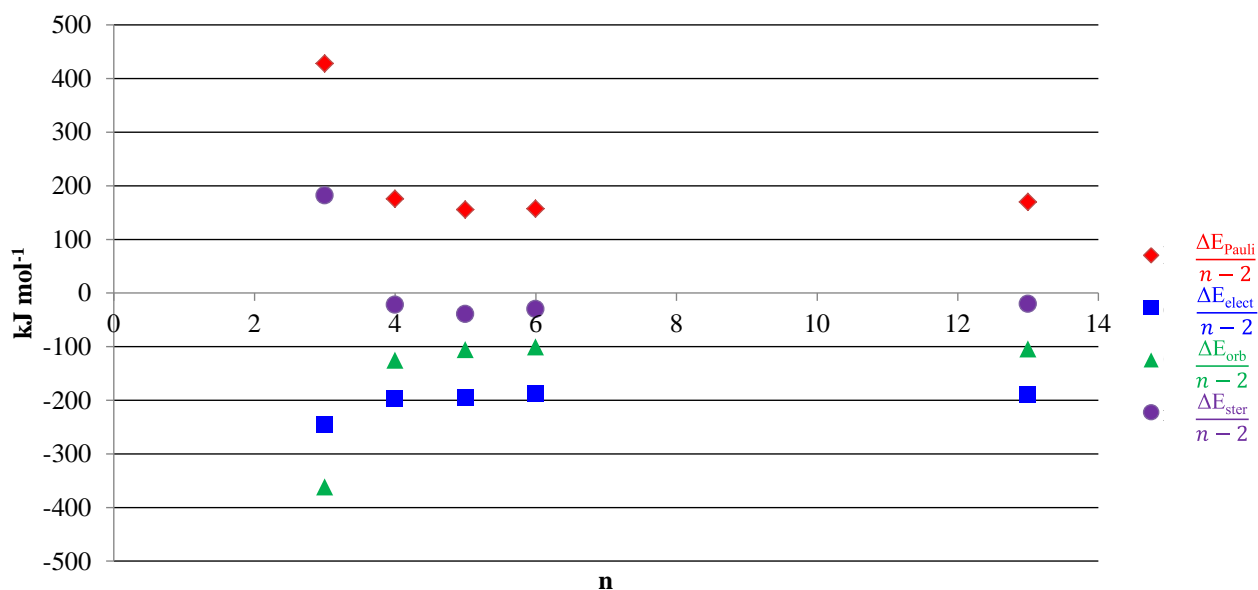


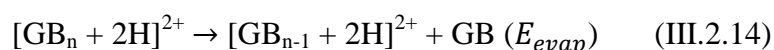
Figure III.2.22. Analyses des contributions énergétiques de l'énergie de cohésion totale pour l'addition d'une GB au sein des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ . Énergies B3LYP-D3/TZP en  $\text{kJ mol}^{-1}$ .

Après avoir identifié les différentes interactions non-covalentes au sein de ces clusters, nous allons maintenant étudier les processus d'élimination d'une GB et de séparation de charge.

### III.2.3.3. Processus de fragmentations

#### III.2.3.3.1. Évaporation d'une glycine bêtaïne neutre

Le processus d'élimination d'un monomère neutre a été observé en spectrométrie de masse en CID et EID (voir III.2.2). Cette voie de fragmentation est représentée par l'équation suivante :



Nous avons considéré que le monomère neutre GB qui se détache en premier du cluster est probablement celui dont l'énergie d'évaporation, dans le cadre *S1*, est le plus faible. Nous l'avons calculé pour chaque GB au sein de chaque cluster  $[\text{GB}_n + 2\text{H}]^{2+}$  (voir Figure III.2.23) et seule l'énergie d'évaporation du GB neutre la plus faible est reportée dans le Tableau III.2.6. Comme observé précédemment, les énergies d'évaporation d'une GB se stabilise à partir de  $n = 4$ . Selon les données M06/6-31+G(d,p) et B3LYP-D3/TZP, l'énergie d'évaporation d'un monomère neutre vaut  $\sim 220 \text{ kJ mol}^{-1}$ . Pour ces mêmes monomères uniquement, nous avons également calculé l'énergie d'évaporation d'un neutre dans le cadre *S2* et une valeur de  $\sim 180 \text{ kJ mol}^{-1}$  a été obtenue.

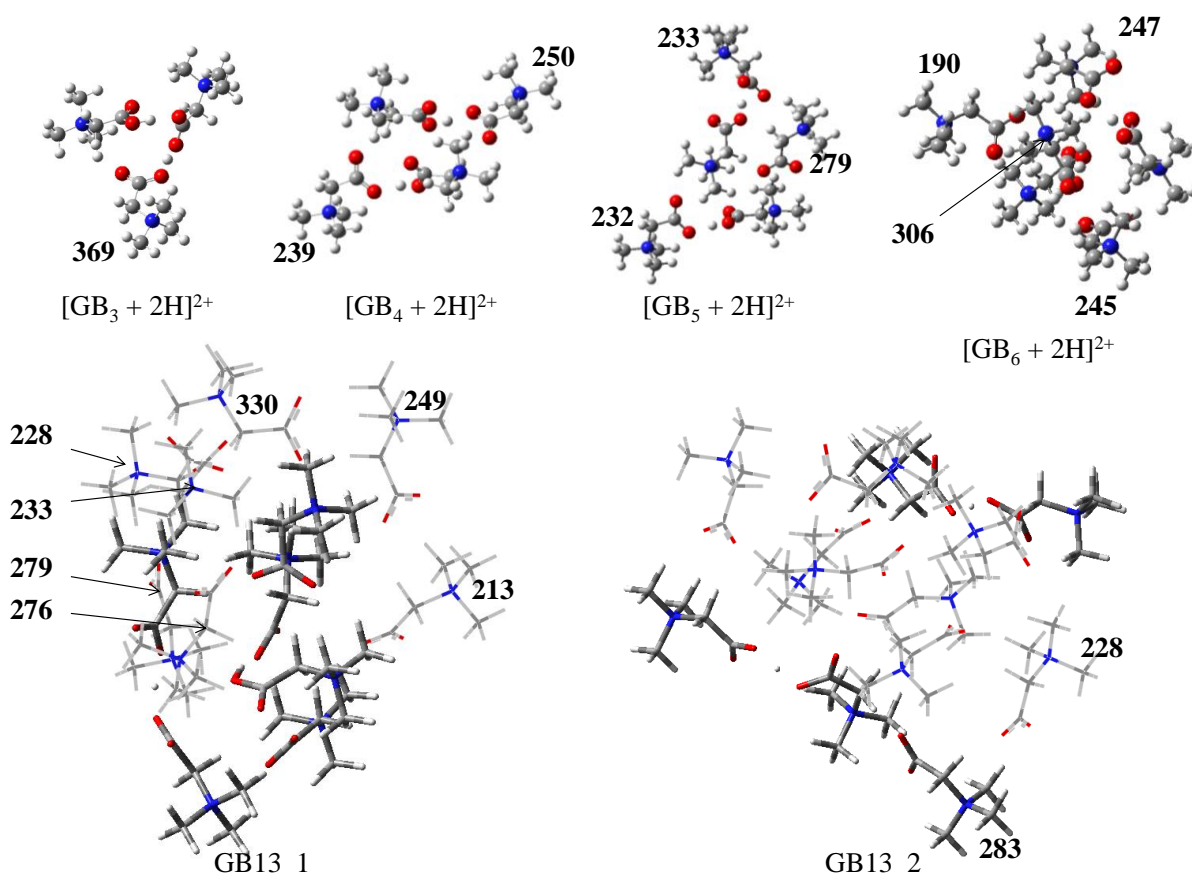


Figure III.2.23. Cohésion entre  $[\text{GB}_n + 2\text{H}]^{2+}$  et un monomère GB. Les valeurs correspondent à des énergies de cohésion au sein du cluster ( $\text{kJ mol}^{-1}$ ).

Tableau III.2.6. Énergies minimales d'évaporation d'un neutre ( $\text{kJ mol}^{-1}$ ).\*

$[\text{GB}_n + 2\text{H}]^{2+}$	$E_{\text{evap}}$	
	M06/6-31+G(d,p)	B3LYP-D3/TZP
$[\text{GB}_2 + 2\text{H}]^{2+}$	/	/
$[\text{GB}_3 + 2\text{H}]^{2+}$	369	363
$[\text{GB}_4 + 2\text{H}]^{2+}$	239 (218)	244
$[\text{GB}_5 + 2\text{H}]^{2+}$	232 (207)	237
$[\text{GB}_6 + 2\text{H}]^{2+}$	190 (144)	191
$[\text{GB}_{13} + 2\text{H}]^{2+}$		
G13_1 <sup>#</sup>	213 (169)	216
G13_2 <sup>#</sup>	228 (188)	217

\* Les valeurs X (Y) correspondent aux approximations *S1* et *S2*, respectivement. <sup>#</sup> Géométries M06/6-31G(d,p\*).

Notez que l'addition d'une GB apporte en moyenne  $\sim -180 \text{ kJ mol}^{-1}$  d'énergie de cohésion au cluster (Tableau III.2.5), tandis que la cohésion d'une GB au cluster vaut  $\sim -220 \text{ kJ mol}^{-1}$  (Tableau III.2.6). Cette différence peut être attribuée au fait que ces deux grandeurs sont calculées de deux façons différentes puisque dans le premier cas nous avons inclus en plus les

énergies d'interactions des deux  $[\text{GB} + \text{H}]^+$ . Mais pour mieux comprendre cette différence, nous avons décomposé l'énergie d'évaporation d'une GB neutre suivant l'équation III.2.13. Les données obtenues sont reportés dans le graphique ci-dessous. En comparant les différentes contributions dans les Figures III.2.22 et III.2.24 (attention au signe qui est adapté aux types de réactions étudiées), on constate que les contributions énergétiques pour la répulsion de Pauli et les interactions orbitales sont peu modifiées. Dans la Figure III.2.24, le cluster  $[\text{GB}_6 + 2\text{H}]^{2+}$  ne suit pas la tendance donnée par les autres clusters, fait qui n'a pas été observé dans la Figure III.2.22.  $[\text{GB}_6 + 2\text{H}]^{2+}$  a directement été extrait de GB13\_1 et non pas par notre stratégie d'exploration. N'étant probablement pas le cluster le plus stable, cela pourrait expliquer qu'il ne suit pas la tendance. Cette différence est compensée dans la Figure III.2.22 car nous avons considéré une valeur moyenne.

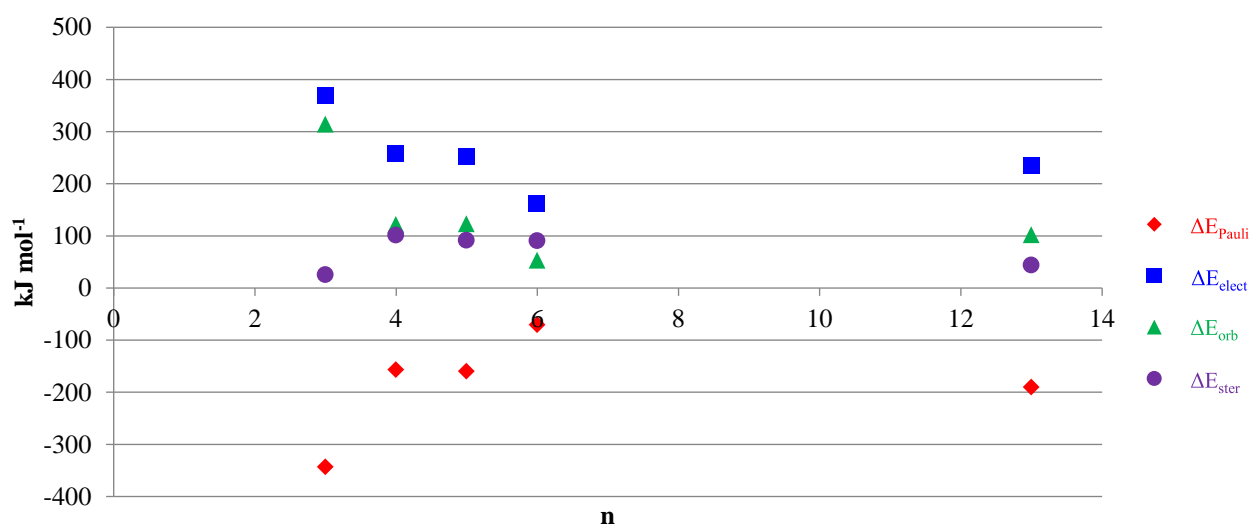
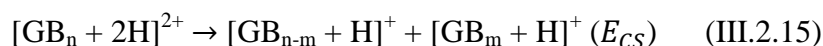


Figure III.2.24. Analyse des contributions énergétiques de l'énergie de cohésion pour l'évaporation d'un neutre. Énergies B3LYP-D3/TZP en  $\text{kJ mol}^{-1}$ .

### III.2.3.3.2. Explosion coulombienne des clusters $[\text{GB}_n + 2\text{H}]^{2+}$

L'équation qui régit ce processus de dissociation est la suivante :



$E_{CS}$  négatif signifie que la réaction est exothermique et donc que le cluster  $[\text{GB}_n + 2\text{H}]^{2+}$  n'existe pas et se dissocie préférentiellement en  $[\text{GB}_{n-m} + \text{H}]^+ + [\text{GB}_m + \text{H}]^+$ . Nous avons adopté une stratégie similaire à celle de l'évaporation d'un neutre. En effet, nous avons testé dans le cadre *SI* différentes possibilités pour l'explosion coulombiennes pour chaque cluster

$[\text{GB}_n + 2\text{H}]^{2+}$  (énergie associée à la séparation de  $[\text{GB}_{n-m} + \text{H}]^+$  et  $[\text{GB}_m + \text{H}]^+$  au sein du cluster). Un exemple pour  $[\text{GB}_{13} + 2\text{H}]^{2+}$  figure dans le Tableau III.2.7. Ensuite, nous avons calculé les énergies de liaison entre les clusters  $[\text{GB}_{n-m} + \text{H}]^+$  et  $[\text{GB}_m + \text{H}]^+$  dans le cas de la voie la plus favorable associée à ce processus de séparation de charge.

Tableau III.2.7. Explosion coulombienne de  $[\text{GB}_{13} + 2\text{H}]^{2+}$  et énergies associées ( $\text{kJ mol}^{-1}$ ).<sup>#</sup>

$[\text{GB}_{n-m} + \text{H}]^+ + [\text{GB}_m + \text{H}]^+$	$E_{CS}$
$[\text{GB}_{12} + \text{H}]^+ + [\text{GB} + \text{H}]^+$	293
$[\text{GB}_{11} + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	282
$[\text{GB}_{10} + \text{H}]^+ + [\text{GB}_3 + \text{H}]^+$	315
$[\text{GB}_9 + \text{H}]^+ + [\text{GB}_4 + \text{H}]^+$	384
$[\text{GB}_7 + \text{H}]^+ + [\text{GB}_6 + \text{H}]^+$	644

<sup>#</sup> Géométries M06/6-31G(d,p\*).

Parmi toutes les combinaisons testées pour les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ , l'énergie de séparation de charge, dans le cadre *S1*, est la plus faible pour  $m = 2$ . Par la suite, seule les données pour  $m = 2$  ont été présentées. Les tendances énergétique entre B3LYP-D3/TZP et M06/6-31+G(d,p) sont en accord (Tableau III.2.8). On constate que la séparation de charge est significativement moins couteuse en énergie pour les petits clusters. En effet, plus  $n$  augmente, plus  $E_{CS}$  augmente et tend vers des valeurs positives.  $E_{CS}$  pour  $[\text{GB}_2 + 2\text{H}]^{2+}$  et  $[\text{GB}_3 + 2\text{H}]^{2+}$  est négative, ce qui implique qu'ils ne sont pas stable en l'état de cluster comme nous l'avons déjà montré précédemment. Dans le cadre *S2*, on constate que  $[\text{GB}_4 + 2\text{H}]^{2+}$  et  $[\text{GB}_6 + 2\text{H}]^{2+}$  sont également plus stables dans leur état dissocié que sous forme de cluster dû à la pénalité énergétique. L'énergie de séparation de charge associée à  $[\text{GB}_5 + 2\text{H}]^{2+}$  est faible ( $\sim 20 \text{ kJ mol}^{-1}$ ). Finalement, pour qu'il y ait séparation de charge au sein des clusters  $[\text{GB}_{13} + 2\text{H}]^{2+}$  il est nécessaire d'apporter de l'énergie au système car ce processus ne se fera pas spontanément comme pour les clusters de plus grandes tailles.

Tableau III.2.8. Énergies minimales pour la séparation des charges ( $\text{kJ mol}^{-1}$ ).<sup>\*</sup>

$[\text{GB}_n + 2\text{H}]^{2+}$	Dissociation	$E_{CS}$	
		M06/6-31+G(d,p)	B3LYP-D3/TZP
$[\text{GB}_2 + 2\text{H}]^{2+}$	$[\text{GB} + \text{H}]^+ + [\text{GB} + \text{H}]^+$	-97 (-104)	-89
$[\text{GB}_3 + 2\text{H}]^{2+}$	$[\text{GB} + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	-22 (/)	-19
$[\text{GB}_4 + 2\text{H}]^{2+}$	$[\text{GB}_2 + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	59 (-10)	60
$[\text{GB}_5 + 2\text{H}]^{2+}$	$[\text{GB}_3 + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	57 (19)	63
$[\text{GB}_6 + 2\text{H}]^{2+}$	$[\text{GB}_4 + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	134 (-43)	135
$[\text{GB}_{13} + 2\text{H}]^{2+}$	$[\text{GB}_{11} + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$		
G13_1 <sup>#</sup>		282 (142)	282
G13_2 <sup>#</sup>		155 (108)	166

<sup>\*</sup>Les valeurs X(Y) correspondent aux approximations *S1* et *S2* respectivement. <sup>#</sup> Géométries M06/6-31G(d,p\*).

Contrairement à l'addition d'une GB et à son évaporation, les différentes contributions énergétiques obtenues (Pauli, électrostatique et orbitaire) pour l'interaction entre  $[\text{GB}_{n-m} + \text{H}]^+$  et  $[\text{GB}_m + \text{H}]^+$  ne tendent pas vers des valeurs limites (Figure III.2.25). Les interactions orbitales et électrostatiques augmentent considérablement avec  $n$  renforçant ainsi l'interaction entre  $[\text{GB}_{n-m} + \text{H}]^+$  et  $[\text{GB}_m + \text{H}]^+$  au sein du cluster. En parallèle la répulsion de Pauli augmente également mais ne permet de compenser les interactions orbitales et électrostatiques pour des  $n$  grands.

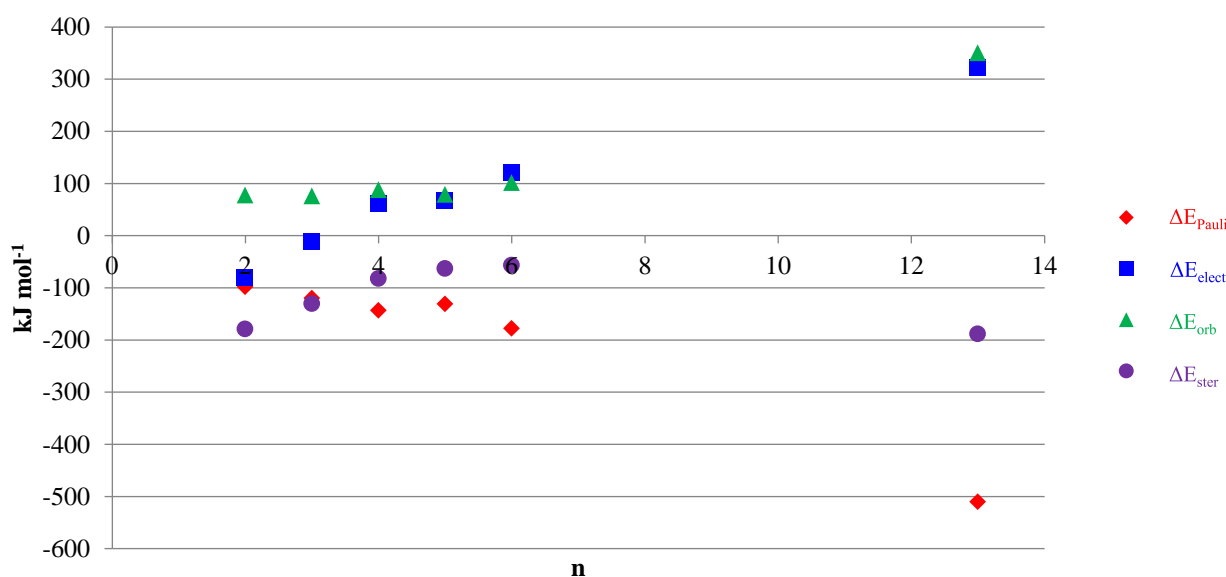


Figure III.2.25. Analyse des contributions énergétiques de l'énergie de cohésion pour l'explosion coulombienne des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ . Énergies B3LYP-D3/TZP en kJ mol<sup>-1</sup>.

Afin de comparer avec la littérature, nous avons calculé les énergies de cohésion entre  $[\text{GB}_{n-m} + \text{H}]^+$  et  $[\text{GB}_m + \text{H}]^+$  avec  $n = 2, 3, 4$  et  $5$  ainsi que le terme de répulsion stérique pour les clusters possédant le motif D (Tableau III.2.9). D'abord, nous avons montré dans la partie II.2.3.1 que les clusters possédant ce motif sont hauts en énergies, le motif (C) étant favorisé. De plus, la cohésion au sein de ces clusters pour les deux fragments chargés n'est pas possible compte tenu des valeurs négatives de  $E_{\text{CS}}$ . La répulsion stérique est également très importante au sein de ces clusters comme l'attestent les valeurs énergétiques. Contrairement à l'hypothèse de Feketeová, nos données montrent que le motif (D) ne peut pas être le motif privilégié au sein des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ .

Tableau III.2.9. Énergies minimales d'évaporation d'un neutre pour les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  possédant le motif D ( $\text{kJ mol}^{-1}$ ).\*

$[\text{GB}_n + 2\text{H}]^{2+}$	Dissociation	$E_{CS}$		$\Delta E_{ster}$
		M06/6-31+G(d,p)	B3LYP-D3/TZP	
$[\text{GB}_2 + 2\text{H}]^{2+}$	$[\text{GB} + \text{H}]^+ + [\text{GB} + \text{H}]^+$	-97	-89	-179
$[\text{GB}_3 + 2\text{H}]^{2+}$	$[\text{GB} + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	-57	-49	-147
$[\text{GB}_4 + 2\text{H}]^{2+}$	$[\text{GB}_2 + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	-33	-25	-127
$[\text{GB}_5 + 2\text{H}]^{2+}$	$[\text{GB}_3 + \text{H}]^+ + [\text{GB}_2 + \text{H}]^+$	-21	-13	-110

### III.2.3.3.3 Bilan

Nous avons reporté dans la Figure III.2.26 les énergies de cohésion et de liaison obtenues en M06/6-31+G(d,p) pour les voies de fragmentations associées à l'évaporation d'un neutre et à la séparation des charges pour les différentes tailles de clusters étudiées. Le comportement de  $[\text{GB}_n + 2\text{H}]^{2+}$  est similaire à celui des clusters  $[\text{SO}_4^{2-}(\text{H}_2\text{O})_n]$ ,  $[\text{Zn}^{2+}(\text{H}_2\text{O})_n]$  et  $[\text{LE}_n]^{2+}$  discutés dans la partie III.2.1.2.1. En effet, l'augmentation de  $n$  privilégie l'élimination d'un neutre par rapport à la séparation de charge.

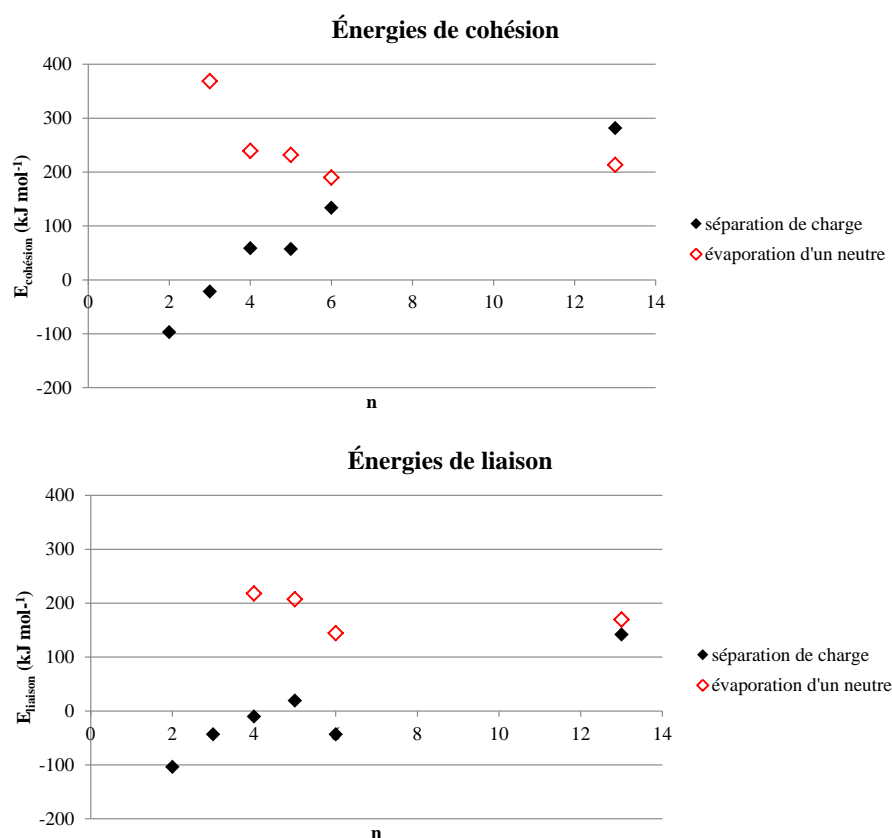


Figure III.2.26. Voies de fragmentation privilégiées pour les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ . Données M06/6-31+G(d,p).

Pour les petits clusters, la séparation de charge est le processus le moins coûteux énergétiquement par rapport à l'évaporation d'un neutre, mais ils deviennent compétitifs en augmentant la taille des clusters. Cela permet d'expliquer que ces deux processus soient observés expérimentalement en EID à partir des mêmes énergies d'électrons. Comme mentionné précédemment, les valeurs négatives pour  $E_{CS}$  ainsi que ces faibles valeurs permettent de justifier que les petits clusters ne soient pas détectés expérimentalement. De plus, compte tenu des valeurs énergétiques théoriques obtenues pour  $[GB_{13} + 2H]^{2+}$ , nous pouvons noter que la cohésion entre  $[GB_{11} + H]^+$  et  $[GB_2 + H]^+$  est la plus faible si les deux protons sont distants ( $155 \text{ kJ mol}^{-1}$  vs  $228 \text{ kJ mol}^{-1}$ ) et que celle du monomère neutre est plus faible si les protons sont proches ( $213$  vs  $282 \text{ kJ mol}^{-1}$ ). Dans le cadre *S2*, la séparation de charge est la voie favorisée dans les deux situations ( $E_{CS} = 142 \text{ kJ mol}^{-1}$  vs  $E_{evap} = 169 \text{ kJ mol}^{-1}$  pour GB13\_1 et  $E_{CS} = 108 \text{ kJ mol}^{-1}$  vs  $E_{evap} = 188 \text{ kJ mol}^{-1}$  pour GB13\_2). Étant donné que GB13\_1 et GB13\_2 sont proches en enthalpies libres et en considérant l'énergie (i) de cohésion entre les différents partenaires et (ii) des différents processus ; nous pouvons supposer que l'existence seule de GB13\_1 permet d'expliquer l'existence de ces deux processus mais aussi que la coexistence de ces structures permet d'expliquer l'existence de ces deux voies de fragmentations puisqu'avec GB13\_2 la séparation de charge est favorable.

### III.2.4. Conclusions et Perspectives

L'utilisation (i) d'AMOEBA pour générer les structures et de (ii) PM6 pour perfectionner les géométries et les énergies permet d'explorer de manière fiable la surface d'énergie potentielle des clusters  $[GB_n + 2H]^{2+}$ . Bien qu'ils soient très flexibles, les clusters les plus stables présentent des structures locales et globales que nous avons identifiées pour toutes les tailles de clusters. Le motif (C) est le motif privilégié adopté par ces clusters. Le cluster réel  $[GB_{13} + 2H]^{2+}$  possède deux types de structures de basses énergies, l'une dont les protons sont proches et l'autre avec les protons éloignés. La structure la plus stable identifiée pour le cluster réel  $[GB_{21} + 2H]^{2+}$  a ces deux protons proches. L'analyse des interactions non covalentes nous a permis de mettre en évidence l'existence d'interactions que nous ne soupçonnions pas initialement à savoir les liaisons hydrogène  $OCO \cdots HC_{CH_2/methyl}$  et interactions de type  $CH \cdots CH$  qui contribuent à la stabilisation des clusters de grandes tailles. Ainsi, il existe non



pas « au plus » mais « au minimum » deux liaisons hydrogène au sein des clusters  $[\text{GB}_n + 2\text{H}]^{2+}$ .

Les deux voies de fragmentations (séparation des charges et élimination) sont compétitives pour des clusters de grandes tailles mais l'explosion coulombienne est (très) favorable pour les clusters  $[\text{GB}_n + 2\text{H}]^{2+}$  de petites tailles ce qui explique d'ailleurs leur absence dans les spectres de masse expérimentaux.

De ces résultats d'autres questions se posent :

- Existent-ils des barrières pour les processus d'explosion coulombienne et d'évaporation d'un GB ? Si oui, sont-elles hautes ?
- Comme mentionné dans la partie III.2.1.2.2, il existe d'autres voies de fragmentations à savoir (i) l'élimination de  $\text{H}^\bullet$  avec un ou plusieurs GB, et (ii) le clivage des liaisons, qui seraient intéressants d'étudier. Sont-elles énergétiquement plus favorable ? Pour le cas (ii), où va l'électron ? Et qu'elle est le mécanisme de fragmentation dans le cas des clivages des liaisons ?
- Pourquoi est-ce que cette voie de dissociation  $[\text{GB}_{21} + 2\text{H}]^{2+} + \text{e}^- \rightarrow \{[\text{GB}_{21} + 2\text{H}]^+ \}^* \rightarrow [\text{GB}_{21-m} + \text{H}]^+ + m\text{GB} + \text{H}^\bullet$  est privilégiée pour des électrons de faible énergie alors que  $[\text{GB}_{21} + 2\text{H}]^{2+} \rightarrow [\text{GB}_{21-m} + 2\text{H}]^{2+} + m\text{GB}$  nécessite l'utilisation d'électron énergétiquement plus haut?

## III.2.5. Bibliographie

- [1] Cooper, T. E.; Carl, D. R.; Armentrout, P. B. *J. Phys. Chem. A* **2009**, *113*, 13727-13741
- [2] Cooper, T. E.; Armentrout, P. B. *J. Phys. Chem. A* **2009**, *113*, 13742-13751.
- [3] Cooper, T. E.; O'Brien, J. T.; Williams, E. R.; Armentrout, P. B. *J. Phys. Chem. A* **2010**, *114*, 12646-12655.
- [4] Blades, A. T.; Kebarle, P. *J. Am. Chem. Soc.* **1994**, *116*, 10761-10766.
- [5] Wong, R. L.; Williams, E. R. *J. Phys. Chem. A* **2003**, *107*, 10976-10983.
- [6] McQuinn, K.; Hof, F.; McIndoe, J. S. *Chem. Commun.* **2007**, 4099-4101.
- [7] Pape, J. McQuinn, K.; Hof, F.; McIndoe, J. S. *Mew J. Chem.* **2011**, *35*, 1582-1587.
- [8] McQuinn, K.; Hof, F.; McIndoe, J. S.; Chen, X.; Wu, G.; Stace, A. J. *Chem. Commun.* **2009**, 4088-4090.
- [9] Wu, R.; McMahon, T. B. *J. Am. Chem. Soc.* **2007**, *129*, 4864-4865.
- [10] Atkins, C. G.; Rajabi, K.; Gillis, E. A. L.; Fridgen, T. D. *J. Phys. Chem. A* **2008**, *112*, 10220-10225.
- [11] Rajabi, K.; Fridgen, T. D. *J. Phys. Chem. A* **2008**, *112*, 23-30.
- [12] Wu, R.; Marta, R. A.; Martens, J. K.; Eldridge, K. R.; McMahon, T. B. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1651-1659.
- [13] Kong, X. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 422-426.
- [14] Raspopov, S. A.; McMahon, T. B. *J. Mass Spectrom.* **2005**, *40*, 1536-1545.
- [15] Vékey, K.; Czira, G. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 783-787.
- [16] Simon, A.; McMahon, T. B. *Int. J. Mass Spectrom.* **2006**, *255-256*, 301-311.
- [17] Armentrout, P. B.; Heaton, A. L.; Ye, S. J. *J. Phys. Chem. A* **2011**, *115*, 11144-11155.
- [18] Price, W. D.; Schnier, P. D.; Williams, E. R. *J. Phys. Chem. B* **1997**, *101*, 664-673.
- [19] Milosavljević, A. R.; Cerovski, V. Z.; Canon, F.; Nahon, L.; Giuliani, A. *Angew. Chem. Int. Ed.* **2013**, *52*, 7286-7290.
- [20] Milosavljević, A. R.; Cerovski, V. Z.; Ranković, M. L.; Canon, F.; Nahon, L.; Giuliani, A. *Eur. Phys. J. D* **2014**, *68*, 68-73.
- [21] Lee, S. W.; Beauchamp, J. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 347-351.
- [22] Jurchen, J. C.; Garcia, D. E.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1373-1386.
- [23] Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1998**, *120*, 3474-3484.
- [24] Feketeová, L.; O'Hair, R. A. J. *Chem. Commun.* **2008**, 4942-4944.
- [25] Feketeová, L.; Khairallah, G. N.; O'Hair, R. A. J. *Eur. J. Mass Spectrom.* **2008**, *14*, 107-110.
- [26] Feketeová, L.; O'Hair, R. A. J. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3259-3263.
- [27] Yoo, E. J. H.; Feketeová, L.; Khairallah, G. N.; O'Hair, R. A. J. *J. Phys. Chem. A* **2011**, *115*, 4179-4185.
- [28] Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J. P.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.



# Chapitre IV : Effets de l'auto-solvatation sur les propriétés électroniques de peptides radical cations



Dans le chapitre précédent, nous avons étudié les modifications des propriétés thermochimiques et structurales d'acides aminés acétylés, induites par la présence explicite de quelques molécules de H<sub>2</sub>O. Dans ce chapitre, nous proposons d'étudier les propriétés électroniques intrinsèques de peptides cationiques riches en hydrogène après la capture d'un électron.

## IV.1. Introduction

La Dissociation par Capture Électronique (ECD)[1,2] et la Dissociation par Transfert Électronique (ETD)[3] sont des méthodes de dissociation initiées par les électrons, que nous avons déjà décrites au chapitre I. Il s'agit d'outils expérimentaux utiles pour l'analyse protéomique. Le peptide multichargé et l'électron vont se recombiner formant ainsi un intermédiaire radical cation dont les voies de fragmentations sont différentes de celles du peptide non réduit (Figure IV.1). En effet, la rupture de la liaison N-C<sub>α</sub> conduit aux fragments de type c et z qui sont spécifiques à l'ETD et l'ECD. Ces deux techniques ont plusieurs avantages dont celui de préserver les modifications post-traductionnelles labiles des peptides.

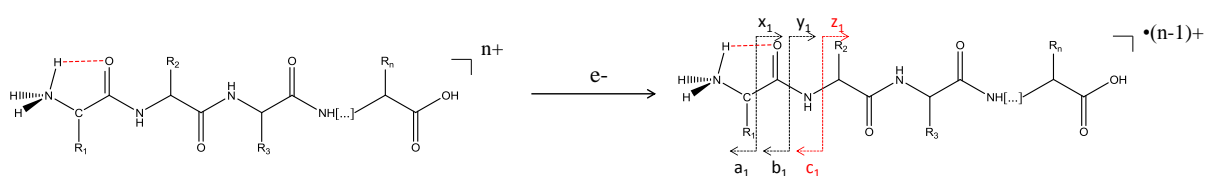


Figure IV.1. Capture électronique d'un peptide multichargé et nomenclature des fragments peptidiques issus de l'ExD.

Etant donné que les mécanismes de fragmentations de ces intermédiaires sont toujours en débats, il est difficile d'établir les relations associant les fragments à la structure primaire du peptide à partir des spectres de fragmentations ECD et ETD. Cependant, les mécanismes proposés se rejoignent sur le fait que les voies de fragmentations sont probablement de type radicalaires. Deux principaux mécanismes ont été postulés à savoir le mécanisme de Cornell[1] proposé par McLafferty et celui d'Utah-Washington[4,5] proposé par Tureček et Simons. La Figure IV.2 donne les principales caractéristiques de ces deux voies de fragmentations. Dans le mécanisme de Cornell, la capture de l'électron se fait par un site chargé et l'intermédiaire radical cation formé est dans son état électronique fondamental. Ensuite, un transfert couplé proton-électron a lieu vers la  $\pi_{CO}^*$ . [6-9] Dans le mécanisme

d'Utah-Washington, l'électron se localise dans la  $\pi_{CO}^*$  ; l'intermédiaire résultant est une super base (radical anion) existant soit dans l'un de ses états électroniques excités soit à l'état fondamental. À ce stade, deux voies ont été proposées, l'une est une fragmentation directe du radical anion par le clivage de la liaison N-C $_{\alpha}$  tandis que la seconde est une réaction de type acide-base qui va permettre le transfert d'un proton vers la super base formant ainsi le même intermédiaire radical cation que dans le mécanisme de Cornell. D'autres mécanismes ont bien évidemment été proposés (voir le paragraphe suivant), mais les premières étapes correspondent également à la capture de l'électron sur un site chargé ou sur une orbitale  $\pi_{CO}^*$  d'un amide.

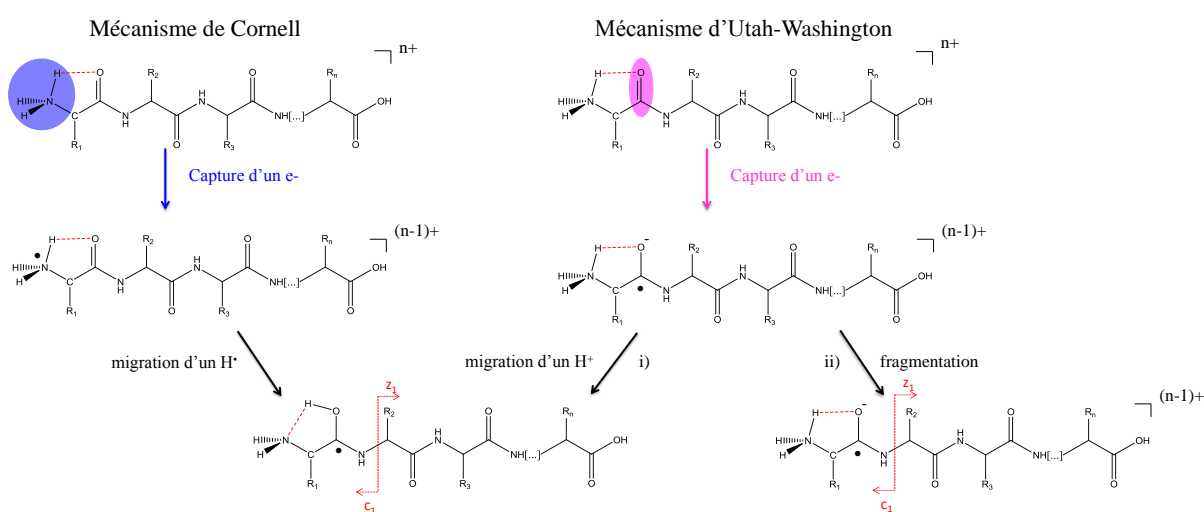


Figure IV.2. Représentation des mécanismes de Cornell et d'Utah-Washington en ECD/ETD.

Un exemple de variante, appelé mécanisme de Uppsala (Figure IV.3) dans lequel le système posséderait une interaction de type  $N_{amide}H \cdots OC$  et l'électron se localiserait dans l'orbitale  $\pi_{CO}^*$ , a également été proposé.[10] L'étape suivante passe par un transfert de proton ce qui mène à la formation d'un site anionique sur l'atome d'azote de l'amide :  $N_{amide}H \cdots OC + e^- \rightarrow N^- + HOC^*$ .

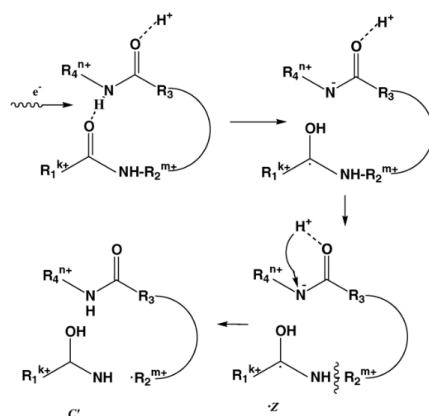


Figure IV.3. Mécanisme Uppsala extrait de la Référence [10].

Il ressort de ces travaux précédents qu'il est important d'identifier la nature des sites pouvant capturer l'électron et d'éventuellement déterminer le ou les sites le(s) plus favorable(s). Pour ce faire, il convient de connaître la structure électronique de l'intermédiaire radical cation dans son état fondamental et dans ses excités au moment de la capture de l'électron. Pour cela une description précise de ces états est nécessaire et il faut donc effectuer un banc d'essai théorique pour identifier une méthode précise et rapide. (cfr section IV.3 ci-dessous).

Expérimentalement, il a été montré que la substitution d'un L-acide aminé par un D-acide aminé modifie les intensités relatives des ions fragments.[11-13] Ainsi, la substitution de la glycine par L- et D-Tryptophane dans la gonadolibérine, noté GnRH (= pyro-Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly), a provoqué une modification des intensités relatives des ions fragments  $z^{+}$  et  $(z + H)^{+}$ , modification qui ne sont pas uniquement localisées sur les fragments provenant des clivages des liaisons situées à proximité de ce résidu substitué.[13] Par ailleurs, Chamot-Rooke et son équipe ont montré que l'intensité des fragments est corrélée à la taille du peptide étudiée.[14] Par exemple, les fragments c/z seront moins abondants pour la dissociation induite par les électrons pour un hexapeptide que pour un dodecapeptide. Cela montre que la conformation et la nature des résidus acides aminés de l'intermédiaire radical cation influencent la dissociation ExD. De plus, les différences mécanistiques proposées dans les mécanismes de fragmentation (Utah-Washington et Uppsala notamment) suggèrent que la nature des liaisons hydrogène liées aux sites de capture de l'électron semble également jouer un rôle déterminant.

Dans ce chapitre, nous avons pour objectif de déterminer s'il existe des influences dues à la présence d'une ou plusieurs liaisons hydrogène autour des sites chargés sur la structure



électronique des états fondamentaux et excités de petits peptides possédant au maximum deux ammoniums. Nous cherchons à obtenir la description la plus précise possible de la structure électronique des radicaux cations intermédiaires. Dans un premier temps, nous avons sélectionné un jeu limité de méthodes de calculs permettant de décrire correctement la structure électronique du système d'intérêt. A notre connaissance, une seule étude précédente concerne l'efficacité de certaines fonctionnelles à décrire de façon fiable la structure électronique de l'état fondamental de peptides radicaux cations.[15] Dans un second temps, nous avons analysé l'effet du nombre de liaison hydrogène environnant un ammonium sur la structure électronique.

## IV. 2. Stratégie d'analyse des structures électroniques

Lorsqu'un peptide multichargé capte un électron, il se retrouve dans un état électronique soit fondamental (GS), soit excité (A). Une énergie de recombinaison, notée ER, est associée à ce processus et vaut  $ER(\mathbf{X}) = E(M_n^+) - E(M^{(n-1)+\bullet}(\mathbf{X}))$  tel que  $\mathbf{X} = \text{GS ou A}$  (voir Figure IV.4). Cette ER peut inclure ou non l'énergie de relaxation ( $\Delta E_{\text{rel}}$ ) de la molécule étudiée. Dans le premier (second) cas, l'évolution de la géométrie après réduction est prise en compte ( $ER_{\text{tot}}$ ) (est négligée ( $ER_{\text{vert}}$ )). On parle de réduction verticale pour ce second cas. Dans notre étude, nous avons considéré les deux approches, même si la réduction verticale a été étudiée uniquement pour s'astreindre des problèmes de géométries, tandis que l'évolution de la géométrie est traitée dans le chapitre suivant.

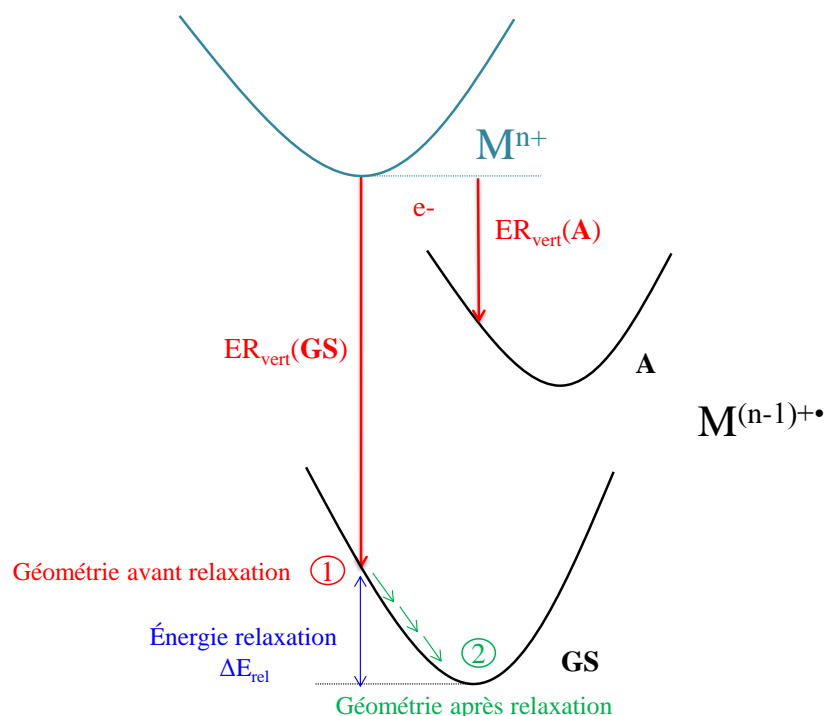


Figure IV.4. Diagramme d'énergie pour la capture d'un électron d'un peptide multichargé  $M^{n+}$  en phase gazeuse.

Afin d'analyser la structure électronique, nous avons tout d'abord étudié les Orbitales Moléculaires (OM) du système dans son état fondamental en DFT puis les états excités associés en TD-DFT. Les états excités sont décrits comme une combinaison d'une ou plusieurs transitions électroniques entre orbitales occupées et virtuelles (OV). Un exemple est donné dans la Figure IV.5 avec  $[\text{GlyLys} + 2\text{H}]^{++}$ . Il ressort que la transition électronique vers le premier état excité résulte d'un grand nombre de composantes orbitales, ces données ne permettent donc pas de visualiser aisément les états excités des peptides radicaux cations. Toutefois, nous avons accès (i) à la différence de densité électronique entre GS et ES1 [16] et (ii) à la densité de spin de ces deux états qui sont toutes deux représentées dans la Figure IV.5. En effet, pour ce dernier cas, les systèmes étudiés sont des systèmes à couche-ouverte ( $S = 2$ ).

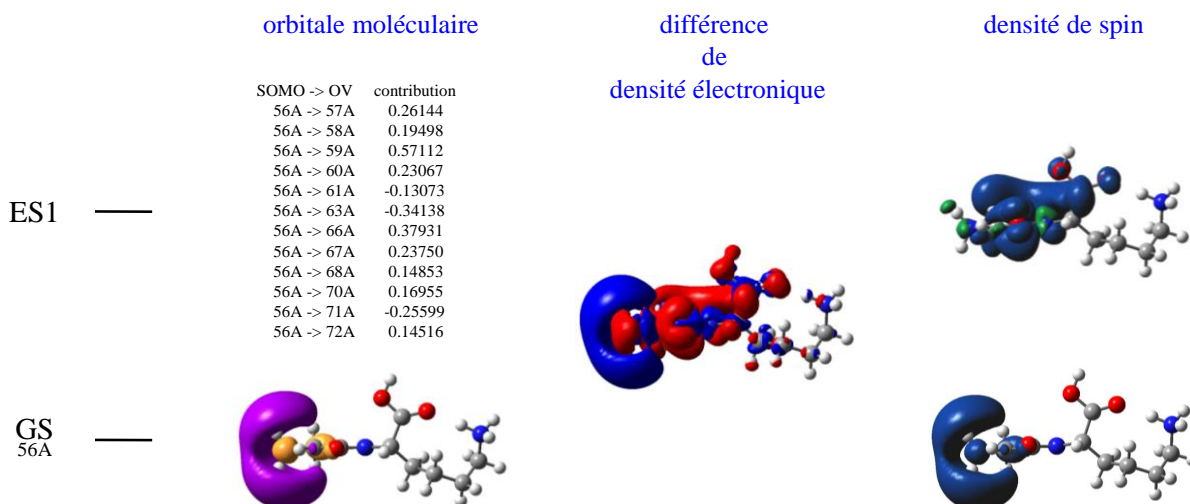


Figure IV.5. Description de la structure électronique du système  $[\text{GlyLys} + 2\text{H}]^{++}$  en fonction des SOMOs (à gauche), de la différence de densité électronique entre ES1 et GS (au centre) et de la densité de spin (à droite).

Grâce à (i), nous pouvons voir que l'électron est localisé sur le N-term de la glycine dans son état fondamental alors qu'après la transition électronique celui-ci se situe sur les deux orbitales  $\pi_{\text{CO}}^*$ . En examinant (ii), on constate que la forme de la densité de spin à l'état fondamental est la même que celle de la SOMO. Pour des raisons de visibilité, nous avons donc préféré travailler avec les densités de spin des différents états électroniques.

### IV.3. Choix des fonctionnelles en DFT et TD-DFT pour l'étude des états fondamental et excités de peptides radicaux cations riches en hydrogène.

Dans la littérature, l'état fondamental voire les états excités des radicaux cations sont souvent calculés à partir de fonctionnelles *habituelles* telles que B3LYP,[17-29] M06-2X[26-29] et CAM-B3LYP[27]. Par exemple, la structure électronique de l'intermédiaire  $[\text{GlyLys} + 2\text{H}]^{++}$  a été calculée au niveau de théorie TD-B3LYP/6-311++G(2d,p) et est représentée dans la Figure IV.6.[17] À l'état fondamental, l'électron est délocalisé sur les deux ammoniums situés de part et d'autre de la molécule. Dans le premier état excité, l'électron est cette fois localisé dans la  $\sigma_{\text{N-H}}^*$  de la lysine. On remarque en sus que l'orbitale de type  $\pi_{\text{CO}}^*$  est située à seulement 0.40 eV au-dessus de l'état fondamental ce qui en fait un état accessible.

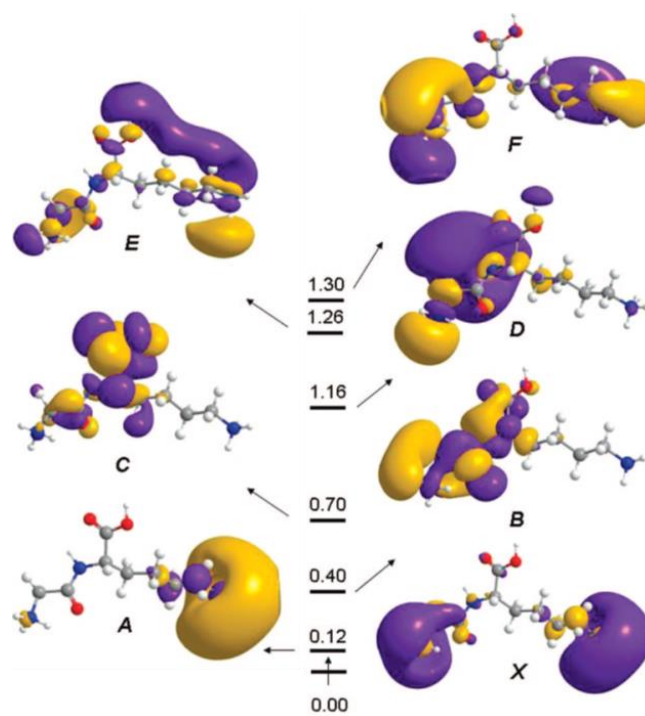


Figure IV.6. Structure électronique de  $[\text{GlyLys} + 2\text{H}]^{++}$  après réduction verticale. Cette image est extraite de la référence [17].

Toutefois, une étude antérieure au laboratoire a permis de montrer que la structure électronique de l'intermédiaire radical cation obtenue après capture d'un électron par l'ion multichargé n'est pas correctement décrite avec les fonctionnelles d'échange-corrélation *conventionnelles*. [15] En effet, après la capture exothermique d'un électron par le système représenté dans la Figure IV.7, une augmentation de la distance entre les deux systèmes devrait diminuer la force de leur interaction jusqu'à ce qu'ils n'interagissent plus. De plus, sans correction de l'erreur d'auto-interaction, la délocalisation de la densité de spin, dans l'état fondamental de cet intermédiaire, est fortement surestimée. Cet effet est donc retrouvé dans la structure électronique de la Figure IV.6. Ceci influence également la description théorique des états excités du radical cation, voies de passage possibles pour expliquer la fragmentation. Seules les fonctionnelles DFT incluant 100 % d'échange exact à grandes distances inter-électroniques permettent de décrire correctement l'état fondamental de ce type de peptide radical cation mono-chargé.

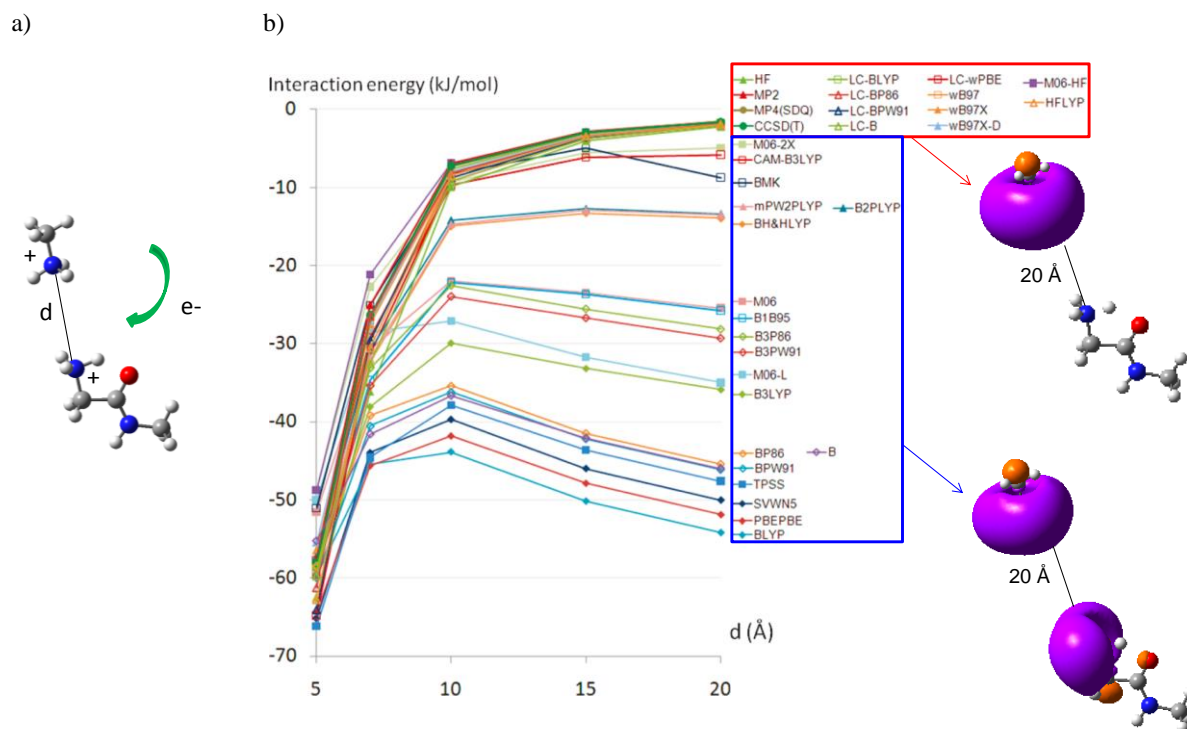


Figure IV.7. a) Structure du système modèle. b) Énergie d'interaction entre les deux systèmes représentés en a) par rapport à la distance  $d$  (à gauche, image extraite de la référence [15]), et les SOMOs associées (à droite).

Toutefois, cette condition est nécessaire mais non suffisante, comme le montre la densité de spin de l'état fondamental obtenu pour une conformation du système  $[\text{GlyLys} + 2\text{H}]^{++}$  en utilisant deux fonctionnelles corrigées à longue-portée (LC-BLYP et  $\omega\text{B97X-D}$ ). En effet, avec LC-BLYP, l'électron est localisé sur le N-term de la glycine tandis que l'état fondamental obtenu avec  $\omega\text{B97X-D}$  correspond au premier état excité LC-BLYP (voir Figures IV.5 et IV.8).

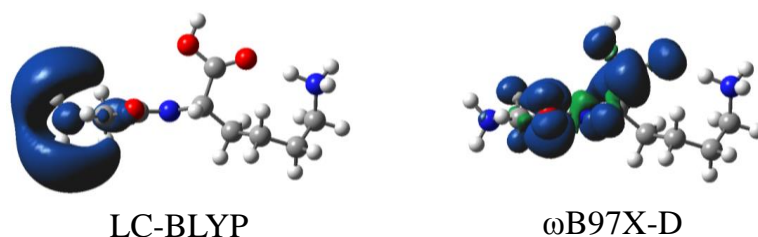


Figure IV.8. Densité de spin d'une conformation de  $[\text{GlyLys} + 2\text{H}]^{++}$  dans son état fondamental obtenue avec LC-BLYP et  $\omega\text{B97X-D}$ .

Pour mieux appréhender ce résultat, nous avons débuté une étude théorique systématique de la structure électronique de l'état fondamental et des états excités de systèmes modèles en

TD-DFT en utilisant la base d'orbitales atomiques 6-311++G(2d,p). Ces résultats, obtenus dans tous les cas avec des fonctionnelles présentant des corrections à longue portée, ont été analysés à la lumière de valeurs de références CASSCF-MRCI. Les systèmes modèles utilisés possèdent un ou deux groupes ammoniums et sont de tailles différentes allant d'un modèle très simple (le méthylammonium) à un hexapeptide ( $[\text{Ala}_5\text{Lys} + 2\text{H}]^{2+}$ ). Ce travail a donné lieu à une collaboration avec le Dr. Denis Jacquemin et le Dr. Émilie Cauët. Il a fait l'objet de la rédaction d'un article scientifique qui a été publié dans *J. Chem. Theory Comput.* et il est reproduit intégralement dans cette thèse. L'ensemble des calculs CASSCF-MRCI ont été effectués par le Dr. Emilie Cauët.

Il est connu que l'efficacité de la méthode de référence CASSCF-MRCI( $n,m$ )<sub>x</sub> dépend fortement du nombre d'électron ( $n$ ) et du nombre d'orbital moléculaire ( $m$ ) composant l'espace actif ainsi que du nombre d'états impliqués ( $x$ ) dans la procédure de calcul. Compte tenu du temps de calcul nécessaire à la réalisation des calculs CASSCF-MRCI( $n,m$ )<sub>x</sub> pour les gros systèmes, différentes tailles de l'espace actif ont été testées sur les petits systèmes afin d'utiliser le plus petit espace actif possible. De plus, nous avons évalué la possibilité de contributions d'excitations multiples. Toujours dans le but de tester sa fiabilité, les énergies relatives des états excités de certains systèmes ont été confrontées à celles obtenues en EOM-CCSD. Les données regroupées dans les Tableaux S1 et S2 appuieront nos arguments sur la validation des calculs CASSCF-MRCI( $n,m$ )<sub>x</sub> dans le corps du texte de l'article. Ces tableaux et d'autres figures utiles à la compréhension du papier seront présentées à la fin de l'article.

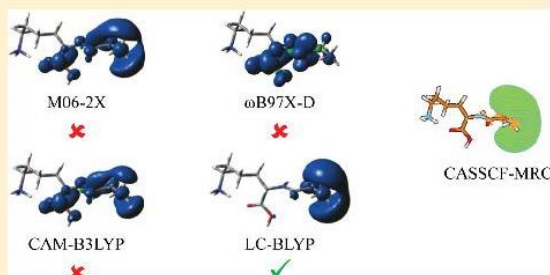


# Benchmarking DFT and TD-DFT Functionals for the Ground and Excited States of Hydrogen-Rich Peptide Radicals

Vanessa Riffet,<sup>†</sup> Denis Jacquemin,<sup>‡,§</sup> Emilie Cauët,<sup>||</sup> and Gilles Frison<sup>\*,†</sup><sup>†</sup>Laboratoire de Chimie Moléculaire, Département de Chimie, Ecole polytechnique and CNRS, 91128 Palaiseau cedex, France<sup>‡</sup>Laboratoire CEISAM, UMR CNRS 6230, Université de Nantes, 2 rue de la Houssinière - BP 92208, 44322 Nantes cedex 3, France<sup>§</sup>Institut Universitaire de France, 103 bd Saint-Michel, F-75005 Paris Cedex 05, France<sup>||</sup>Service de Chimie quantique et Photophysique, Université Libre de Bruxelles, CP160/09, 50 av. F.D. Roosevelt, 1050 Bruxelles, Belgium

## Supporting Information

**ABSTRACT:** We assess the pros and cons of a large panel of DFT exchange-correlation functionals for the prediction of the electronic structure of hydrogen-rich peptide radicals formed after electron attachment on a protonated peptide. Indeed, despite its importance in the understanding of the chemical changes associated with the reduction step, the question of the attachment site of an electron and, more generally, of the reduced species formed in the gas phase through electron-induced dissociation (ExD) processes in mass spectrometry is still a matter of debate. For hydrogen-rich peptide radicals in which several positive groups and low-lying  $\pi^*$  orbitals can capture the incoming electron in ExD, inclusion of full Hartree–Fock exchange at long-range interelectronic distance is a prerequisite for an accurate description of the electronic states, thereby excluding several popular exchange-correlation functionals, e.g., B3LYP, M06-2X, or CAM-B3LYP. However, we show that this condition is not sufficient by comparing the results obtained with asymptotically correct range-separated hybrids (M11, LC-BLYP, LC-BPW91,  $\omega$ B97,  $\omega$ B97X, and  $\omega$ B97X-D) and with reference CASSCF-MRCI and EOM-CCSD calculations. The attenuation parameter  $\omega$  significantly tunes the spin density distribution and the excited states vertical energies. The investigated model structures, ranging from methylammonium to hexapeptide, allow us to obtain a description of the nature and energy of the electronic states, depending on (i) the presence of hydrogen bond(s) around the cationic site(s), (ii) the presence of  $\pi^*$  molecular orbitals (MOs), and (iii) the selected DFT approach. It turns out that, in the present framework, LC-BLYP and  $\omega$ B97 yields the most accurate results.



## INTRODUCTION

Electron capture dissociation (ECD)<sup>1,2</sup> and electron transfer dissociation (ETD)<sup>3</sup> are recent fragmentation techniques that possess strong potential for the analysis of both peptides and proteins.<sup>4–6</sup> In ECD and ETD, multiply protonated peptides are partially reduced by electron capture and transfer, respectively, and are subsequently converted from even-electron closed-shell species, to cation-radical intermediates that fragment mainly through N–C $\alpha$  bond cleavage,<sup>7</sup> leading to c/z fragments<sup>5</sup> that are specific to these methods. The labile post-translational modifications are well-preserved during the fragmentation process,<sup>8–13</sup> which is a great advantage of these methods, compared to more-standard techniques, such as collision-induced dissociation (CID). Since the reaction mechanism remains unclear, efficient empiric models able to predict the fragmentation of peptides are not available, which severely restricts the automatic determination of the peptide and protein primary structure information on the basis of the sole recorded fragmentation's spectrum. To date, several mechanisms have

been proposed, which could be classified by the attachment site of the incoming electron resulting in a charge-reduced cation-radical intermediate. This intermediate can be also described as an hydrogen-rich peptide–cation radical.<sup>14</sup> Several electron attachment sites have been proposed:

- one of the charged sites (N-terminal amine group or a basic amino-acid lysine, arginine, or histidine side chain) of the multiply protonated peptides;<sup>1,15–17</sup>
- a hydrogen bond between backbone carbonyl and nitrogen groups;<sup>18</sup> or
- an amide  $\pi^*$ -orbital,<sup>19–21</sup> producing a charge-stabilized amide anion-radical intermediate in an electronic excited state (ES) but also possibly in the ground state (GS).<sup>22,23</sup>

In all cases, a homolytic N–C $\alpha$  bond cleavage occurs either before or after proton and/or hydrogen migration. Other

Received: April 9, 2014

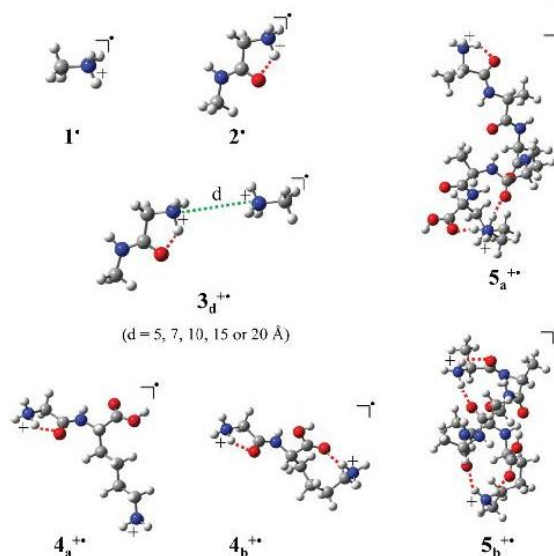
Published: June 26, 2014



fragmentation pathways than N–C $\alpha$  bond cleavage have also been observed experimentally.<sup>24–34</sup>

Understanding the radiation damage process has motivated extensive theoretical and experimental works devoted to the electron attachment in biological molecules. [For recent reviews, see refs 35 and 36.] For positively charged isolated peptides, a model based on *ab initio* calculations has been proposed.<sup>37–39</sup> This model allows one to analyze both the electron attachment and intramolecular electron transfer from ground-state and excited-Rydberg orbitals to the amide  $\pi^*$ -orbital, based on their principal quantum number, size, and energy. In that framework, the accurate description of the electronic structure of charge-reduced peptides in both the GS and ES is a critical prerequisite for understanding the mechanism of fragmentation in ECD and ETD. Previous computational works on related systems mainly used the B3LYP exchange-correlation DFT functional for the GS and its time-dependent density functional theory (TD-B3LYP) framework for the ESs.<sup>22,40–48</sup> Other DFT functionals, such as M06-2X and CAM-B3LYP, have also been applied in recent studies of hydrogen-rich peptide radicals.<sup>23,48–50</sup> In a recent study, we have shown that DFT functionals that do not include full Hartree–Fock exchange ( $X^{HF}$ ) at long-range interelectronic distance are unable to properly describe the GS electronic structure of the charge-reduced peptide cation radicals.<sup>51</sup> Indeed, because of the self-interaction error, conventional DFT functionals yield an overly disperse spin density after addition or removal of an electron from closed-shell systems<sup>52</sup> or show an excessive charge-transfer when describing radical-molecule interactions.<sup>53</sup> Only HF and post-HF methods, DFT functionals with 100%  $X^{HF}$ , and range-separated hybrid (RSH) functionals with correct asymptotic behavior provide an accurate description, in terms of both radical-cation intermolecular interaction energy and localization of spin density. Similar trends have been reported for the prediction of the spin density of organometallic species.<sup>54</sup> An inaccurate characterization of the spin localization could be a major problem for describing electron paramagnetic resonance (EPR) properties, because spin density is an essential ingredient in accurate EPR calculations. Contrary to the GS of hydrogen-rich peptide radicals for which computational methods have been probed,<sup>51</sup> ESs have not been the subject of benchmark studies to determine which method describes them accurately. In addition, the natures and energies of the latter remain, to the best of our knowledge, unexplored with RSHs methods. In a charge-reduced peptide, it can be postulated that the electron is localized on the charged site with the highest electron affinity; however, the energy required to promote the electron to the other cationic site(s) or to an amide  $\pi^*$ -orbital remains unknown. It is well-established that global and range-separated functionals (RSHs) can yield very different TD-DFT results.<sup>55</sup> Most of the numerous existing TD-DFT and *ab initio* benchmark studies focus on the excitations energies of closed-shell molecules including medium-sized<sup>56</sup> to large-sized<sup>57</sup> unsaturated organic molecules or transition-metal complexes.<sup>58</sup>

In the present contribution, a large panel of exchange-correlation DFT functionals is used to investigate both the GS and ESs of radicals in the TD-DFT framework, and the TD-DFT results are benchmarked using reference data obtained with the CASSCF-MRCI wave function method as well as with the equation-of-motion coupled-cluster approach (EOM-CCSD) approach. Several chemical models of increasing complexity are studied (see Figure 1). First, we examine neutral radicals, including a NH<sub>3</sub> group either free (methylammonium 1 $^{\bullet}$ ) or implied in a hydrogen bond to an amide carbonyl group



**Figure 1.** Studied model systems. Intramolecular hydrogen bonds to the ammonium group are depicted with red dashes.

(protonated glycine amide 2 $^{\bullet}$ ). Next, we combine these molecules to build the charge-reduced cation-radical peptide model 3 $d^{\bullet\bullet}$  including two NH<sub>3</sub> groups separated by a fixed distance  $d$  ( $d = 5, 7, 10, 15$ , or  $20$  Å). Finally, we extend our studies to real monoreduced diprotonated peptides. This includes both the dipeptide GlyLys, for which two conformations presenting different intramolecular hydrogen bonds have been computed (models 4 $a^{\bullet\bullet}$  and 4 $b^{\bullet\bullet}$ ),<sup>42</sup> and the hexapeptide Ala<sub>5</sub>Lys (conformation models 5 $a^{\bullet\bullet}$  and 5 $b^{\bullet\bullet}$ ) studied recently by Wodrich and co-workers.<sup>15</sup>

## METHODS

All DFT and TD-DFT calculations have been performed with the Gaussian 09 program.<sup>59</sup> In this work, vertical reduction have been considered to avoid the complexity of considering simultaneously several ES geometries. Therefore, geometries used for all radicals correspond to the optimized structures of the nonreduced molecules. They have been obtained at the M06/6-311++G(d,p) level for 1 $^{\bullet}$  and 2 $^{\bullet}$ , or taken from the literature for 3 $d^{2+}$ ,<sup>51</sup> 4 $a^{2+}$ /4 $b^{2+}$ ,<sup>42</sup> and 5 $a^{2+}$ /5 $b^{2+}$ .<sup>15</sup> A large panel of DFT functionals has been applied for these calculations. They can be classified into three categories: (a) the RSHs CAM-B3LYP,<sup>60</sup> HISSbPBE,<sup>61</sup>  $\omega$ B97X-D,<sup>62</sup>  $\omega$ B97X,<sup>63</sup>  $\omega$ B97,<sup>63</sup> LC-BLYP,<sup>64,65</sup> LC-BPW91,<sup>64,65</sup> and M11;<sup>66</sup> (b) functionals including 100%  $X^{HF}$ , that is, M06-HF,<sup>67,68</sup> and HFLYP;<sup>69,70</sup> and (c) conventional pure and hybrid functionals, such as BLYP,<sup>69–71</sup> B3LYP,<sup>69,72</sup> M06,<sup>73</sup> BH&HLYP,<sup>74</sup> and M06-2X.<sup>73</sup> RSHs present a growing fraction of  $X^{HF}$  with increasing interelectronic separation, whereas the amount of  $X^{HF}$  is constant in all other methods (see Table 1). In all RSHs but HISSbPBE, the increased rate of  $X^{HF}$  is directly governed by the attenuation parameter  $\omega$ , according to the scheme of eq 1:

$$\frac{1}{r_{12}} = \frac{\text{erf}(\omega r_{12})}{r_{12}} + \frac{1 - \text{erf}(\omega r_{12})}{r_{12}} \quad (1)$$

where erf is the error function and  $r_{12}$  is the interelectronic distance ( $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ ). The larger the  $\omega$  value, the faster the



Table 1. Mean Absolute Deviations (in eV) of the Absolute and Relative Vertical Excitation Energies for 1<sup>•</sup> and 2<sup>•</sup>

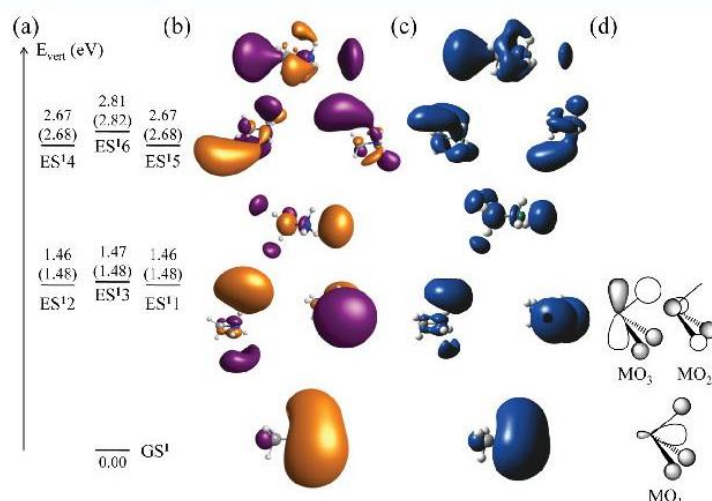
functional	methods		MAD(E) <sup>c</sup>		MAD(ΔE) <sup>d</sup>		MAD <sub>average</sub>
	n (%) <sup>a</sup>	ω <sup>b</sup> (a.u.)	1 <sup>•</sup>	2 <sup>•</sup>	1 <sup>•</sup>	2 <sup>•</sup>	
BLYP	0		0.14	0.44	0.24	0.31	0.28
B3LYP	20		0.10	0.27	0.18	0.20	0.19
M06	27		0.40	0.39	0.30	0.25	0.33
BH&HLYP	50		0.06	0.09	0.10	0.04	0.07
M06-2X	54		0.12	0.10	0.19	0.15	0.14
HISSbPBE	0–60–0	0.84/0.20 <sup>e</sup>	0.14	0.21	0.17	0.07	0.15
CAM-B3LYP	19–65	0.33	0.03	0.06	0.07	0.17	0.08
ωB97X-D	22.2–100	0.20	0.05	0.13	0.11	0.31	0.15
ωB97X	15.77–100	0.30	0.05	0.13	0.10	0.29	0.14
M11	42.8–100	0.25	0.44	0.22	0.25	0.17	0.27
ωB97	0–100	0.30	0.06	0.10	0.08	0.25	0.12
LC-BLYP	0–100	0.47	0.04	0.11	0.07	0.23	0.11
LC-BPW91	0–100	0.47	0.09	0.19	0.16	0.42	0.22
M06-HF	100		0.15	f	0.18	f	0.17
HFLYP <sup>g</sup>	100		0.05	0.10	0.07	0.22	0.11
HF			0.14	0.09	0.10	0.24	0.14
EOM-CCSD			0.01	0.03	0.01	0.07	0.03

<sup>a</sup>Exact exchange percentage for DFT methods. <sup>b</sup>Attenuation parameters for long-range corrected functional. <sup>c</sup>Mean absolute deviations of the TD-DFT vertical excitation energies (MAD(E)), relative to CASSCF-MRCI ones. <sup>d</sup>Mean absolute deviations of the TD-DFT energy gaps between excited states (ESs).  $\Delta E = E(\text{ES}_{i,1}^X) - E(\text{ES}_i^X)$ ;  $X = 1, 2$ ;  $i = 1-5$  for 1<sup>•</sup> and  $i = 1-2$  for 2<sup>•</sup>. <sup>e</sup>ω at short range (ω<sub>SR</sub>)/ω at long range (ω<sub>LR</sub>); the rate of X<sup>HF</sup> is governed by these parameters according to:  $1/r_{12} = [(1 - \text{erf}(\omega_{\text{SR}}r_{12}))/r_{12}] + (\text{erf}(\omega_{\text{LR}}r_{12}))/r_{12} + \{[1 - \text{erf}(\omega_{\text{LR}}r_{12})] - [1 - \text{erf}(\omega_{\text{SR}}r_{12})]\}/r_{12}$ . <sup>f</sup>No data available. <sup>g</sup>Functional including full-X<sup>HF</sup> part (no local and nonlocal exchange functional) and the Lee–Yang–Parr correlation functional. Obtained with G09 keywords and iop: “BLYP iop(3/76=0000010000) iop(3/77=0720008000) iop(3/78=0810010000)”.

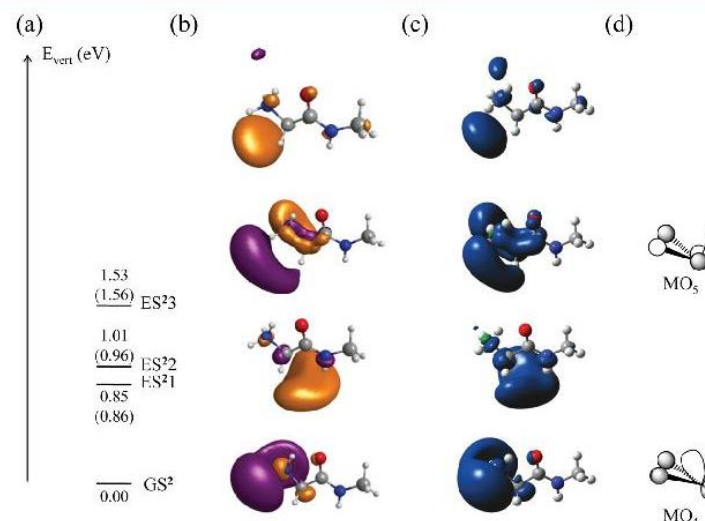
transition from Kohn–Sham exchange (X<sup>KS</sup>) to X<sup>HF</sup>. The change between X<sup>KS</sup> and X<sup>HF</sup> differs for HISSbPBE, which possesses X<sup>HF</sup> at middle range but only X<sup>KS</sup> at short range and long range. It is well-known that the parameter ω can be adjusted to obtain a better correlation between theory and experience, either empirically<sup>75</sup> or following the optimal tuning approach recently developed by Baer and co-workers.<sup>76–80</sup> In this approach, ω is not a universal constant but rather a system-specific range-separation parameter that satisfies to the best possible degree the ionization potential theorem.<sup>81</sup> The optimally tuned range separated hybrid functionals have been shown, for example, to outperform many of the most efficient functionals for excited-state transition energies.<sup>82</sup> For all cases, low-energy ES of hydrogen-rich peptide radicals correspond to the excitation of a single electron. This assumption was confirmed by CASSCF-MRCI and EOM-CCSD calculations (*vide infra*). Consequently, the nature of both the GS and the ESs could be either described by the singly occupied molecular orbital (SOMO) of the corresponding state, or by the spin density distribution. This last description is more convenient for TD-DFT calculations for which, in most cases, several vacant MOs significantly contribute to the description of ESs. Vertical excited-state energies and molecular spin densities of the ground and excited states have been computed using DFT and TD-DFT calculations. All calculations on radicals have been performed with the 6-311++G(2d,p) atomic basis set. Their shapes were visualized applying a contour threshold of 0.001 a.u. Atomic spin densities have been calculated using the natural population analysis (NPA) method.<sup>83</sup>

Because of the need to ensure the quality of our reference data for benchmarking the DFT performances, two high-level *ab initio* methods, including CASSCF-MRCI and EOM-CCSD, have been used. Multiconfiguration *ab initio* calculations on 1<sup>•</sup>, 2<sup>•</sup>, 3<sub>d</sub><sup>•</sup>, and 4<sub>a/b</sub><sup>•</sup> have been performed using the same 6-311++G(2d,p) atomic basis set. These complete active space self-

consistent field (CASSCF)<sup>84</sup> calculations have been performed in which the molecular orbitals playing an active role in the description of the ground and low-lying excited states were optimized. A state-averaging procedure has been used to optimize a common molecular orbital basis set describing the electronic states of the molecules. Main configuration state functions (CSFs) defining the low-lying states have been taken into account in a restricted active orbital space correlating a minimal number of electrons. In order to account for the remaining dynamical correlation, the CASSCF calculations have been followed by internally contracted multireference configuration interaction (MRCI) calculations.<sup>85</sup> All multiconfiguration *ab initio* calculations were performed with the MOLPRO program package.<sup>86</sup> The size of the active orbital space is defined by the standard notation (*n,m*), indicating that *n* electrons are distributed among *m* molecular orbitals, with *x* being the number of states involved in the state averaging procedure. The choice of this active space is a key issue, because of both the required computer time and the sensitivity of the MRCI step to its size. It has been postulated in the literature<sup>42</sup> that all low-lying excited states of hydrogen-rich peptide radicals correspond to excitation of the single electron. To validate this assumption and to exclude the possibility of multielectron excitation, CASSCF-MRCI calculations with an active space including one or five electrons in five orbitals were performed for 3<sub>s</sub><sup>•</sup> (see Table S1 in the Supporting Information (SI)). Equal excitation energies obtained from (1,5)<sub>3</sub> and (5,5)<sub>3</sub> active spaces justify, in the following, the use of active space including only one electron. Similarly, the influence of the number of molecular orbitals in the active space, as well as the number of states involved in the state averaging procedure, has been evaluated for 1<sup>•</sup>, 2<sup>•</sup>, 3<sub>s</sub><sup>•</sup>, and 3<sub>10</sub><sup>•</sup>. In all cases, the first excitation energies are similar, with a larger deviation of 0.04 eV for the second excited state of 2<sup>•</sup> (see Table S1 in the SI). Consequently, minimal active orbital spaces have been used in CASSCF calculations: (1,10)<sub>7</sub> for 1<sup>•</sup> and 2<sup>•</sup>



**Figure 2.** Vertical electronic states of **1\***: (a) CASSCF-MRCI and EOM-CCSD (in bracket) energies, (b) CASSCF-MRCI SOMO, (c) LC-BLYP and TD-LC-BLYP spin density distributions of the ground and excited states, and (d) SALC-MOs from  $\sigma_{N-H}^*$  orbitals in a  $XH_3$ -type molecule ( $X = N$ ) with a pyramidal geometry.



**Figure 3.** Vertical electronic states of **2\***: (a) CASSCF-MRCI and EOM-CCSD (in bracket) energies, (b) CASSCF-MRCI SOMO, (c) LC-BLYP and TD-LC-BLYP spin density distributions of the ground and excited states, and (d) SALC-MOs from  $\sigma_{N-H}^*$  orbitals in a bend  $XH_2$ -type molecule ( $X = N$ ).

and  $(1,3)_3$  for  $3_d^{+*}$  and  $4_{a/b}^{+*}$ . These spaces are defined by the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbitals (LUMO to LUMO+8 for **1\*** and **2\***, and LUMO and LUMO+1 for  $3_d^{+*}$  and  $4_{a/b}^{+*}$ ) in which one electron is distributed to form the CASSCF wave functions. These wave functions provided the smaller multiconfigurational description of the seven lowest states of the **1\*** and **2\*** systems and of the three lowest states for the  $3_d^{+*}$  and  $4_{a/b}^{+*}$ . For  $3_d^{+*}$  and  $4_{a/b}^{+*}$ , a significant amount of computer time is required for variationally optimizing the 33 and 55 closed-shell orbitals, respectively, in addition to the active orbitals. For this reason, we decided to freeze the 16 and 28 lowest occupied molecular orbitals for  $3_d^{+*}$  and  $4_{a/b}^{+*}$ , respectively, arising from the 1s and 2s orbitals of O, C, and N atoms that were optimized at the SCF

level. The state-averaged CASSCF calculations with the minimal active orbital spaces were performed as a preliminary orbital optimization step to prepare for the MRCI calculations. The multireference wave function used in the MRCI calculations exactly corresponds to its CASSCF counterpart. All single and double excitations from the active orbitals to the external orbital space were included in the MRCI expansion whose size reached  $10^6$ ,  $10^6$ ,  $10^7$ , and  $10^8$  contracted CSFs, for **1\***, **2\***,  $3_d^{+*}$ , and  $4_{a/b}^{+*}$ , respectively. This level of calculation, noted CASSCF-MRCI in the following, takes into account both dynamic and static correlation effects of the active electrons providing an accurate representation of the electronic states of the chemical models. All MRCI energies include the Davidson correction for unlinked quadruple excitations.<sup>87</sup> This correction was calculated with a

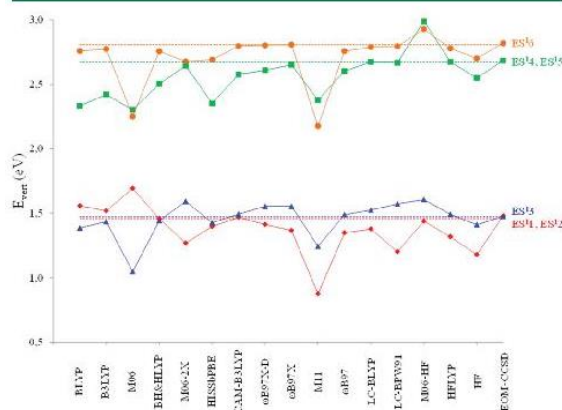


relaxed multireference function as implemented in MOLPRO. The SOMO of each state can be visualized, but atomic spin densities are not available.

To further guarantee the reliability and accuracy of our CASSCF-MRCI reference calculations, relative excited-state energies were evaluated with the equation-of-motion coupled cluster method in the singles and doubles approximation (EOM-CCSD).<sup>88</sup> In this method, the ES wave function is generated from the coupled-cluster singles and doubles reference (CCSD) wave function of the ground state. EOM-CCSD calculations have been performed for  $1^\bullet$ ,  $2^\bullet$ , and  $3_d^{\bullet\bullet}$  with the 6-311++G(2d,p) basis set used for other methods. Because of the well-known sensitivity of the basis set to EOM-CCSD results, comparison have been performed by using the aug-cc-pVTZ basis for  $1^\bullet$  and  $2^\bullet$ . The variations, which remain low (see Table S2 in the SI), confirm the relevance of using the 6-311++G(2d,p) basis set. GS molecular parameters (molecular orbitals and atomic spin density) are computed at the CCSD level of theory. In all cases, ESs result from the excitation of the electron located in the SOMO of the GS. In most cases, several vacant molecular orbitals (MOs) significantly contribute to the description of ESs, which makes difficult to determine the nature of these ESs.

## RESULTS AND DISCUSSION

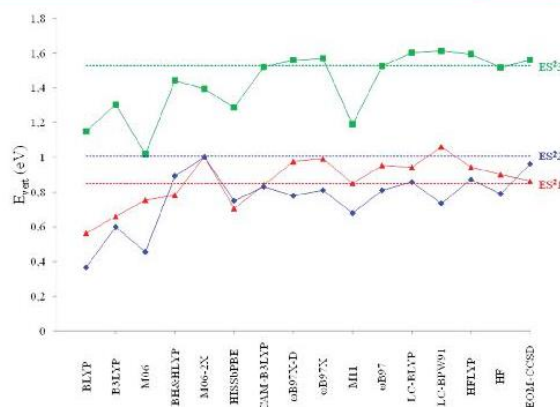
**Benchmarks on Hydrogen-Rich Neutral Radicals.** We started our study with the small neutral models  $1^\bullet$  and  $2^\bullet$  that



**Figure 4.** Vertical excitation energies of  $1^\bullet$  obtained at the TD-DFT and EOM-CCSD levels. The dotted line corresponds to the CASSCF-MRCI reference.

include a single  $\text{NH}_3$  group. For these models, belonging to the  $C_3$  and  $C_s$  point group, respectively, a chemically intuitive description of the electronic states could be achieved. Furthermore, they correspond to models for which conventional DFT should be able to describe the GS accurately, contrary to models including two  $\text{NH}_3$  groups.<sup>51</sup> Fifteen DFT functionals have been used on these models to assess their accuracy, in terms of absolute and relative energies of the first ESs, as well as spin densities.

As expected, for all methods, the single electron is unequivocally located on the charged group for both the GS of  $1^\bullet$  (noted  $\text{GS}^1$ ) and  $2^\bullet$  (noted  $\text{GS}^2$ ). The electronic structures obtained at the CASSCF-MRCI level are represented in Figures 2 ( $1^\bullet$ ) and 3 ( $2^\bullet$ ), respectively. For  $1^\bullet$ , we observed degeneracies of the first ( $\text{ES}^1, \text{ES}^2$ ) and third ( $\text{ES}^4, \text{ES}^5$ ) ESs, because of

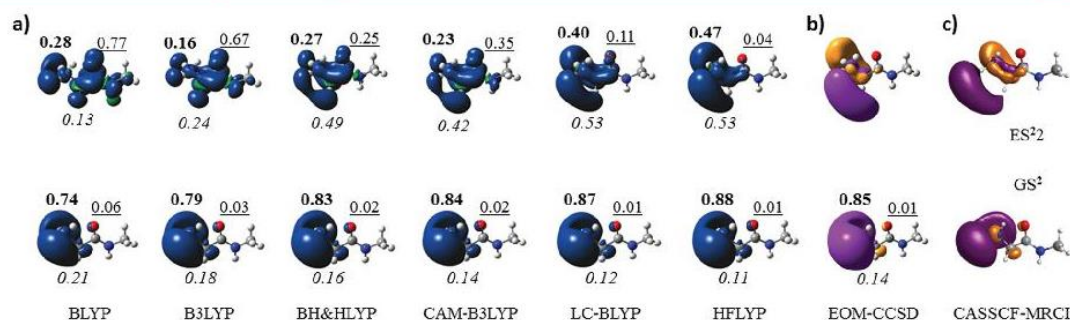


**Figure 5.** Vertical excitation energies of  $2^\bullet$  (see caption of Figure 4 for more details).

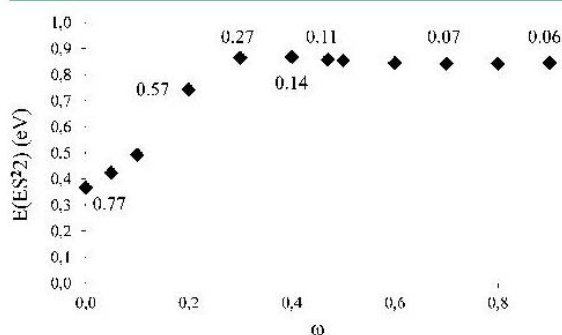
the  $C_3$  symmetry axis. In addition,  $\text{ES}^3$  is close (0.01 eV) to the two former ESs. The shape of the SOMO of the  $\text{GS}^1$ ,  $\text{ES}^1$ , and  $\text{ES}^2$ , is clearly reminiscent of the traditional antibonding molecular orbitals ( $\text{MO}_1$ ,  $\text{MO}_2$ , and  $\text{MO}_3$ ) of  $\text{XH}_3$ -type molecules with a pyramidal geometry.  $\text{GS}^1$  corresponds to the in-phase combination of the three  $\sigma_{\text{N-H}}$  orbitals ( $\text{MO}_1$  in Figure 2), whereas  $\text{ES}^1$  and  $\text{ES}^2$  can be represented by the two other symmetry-adapted linear combinations (SALCs,  $\text{MO}_2$ , and  $\text{MO}_3$ ) of these  $\sigma_{\text{N-H}}$  orbitals. The SOMO in  $\text{ES}^3$ - $\text{ES}^6$  are mainly located on the  $\text{CH}_3$  group, and could be described using both the  $\sigma_{\text{N-C}}$  orbital and the SALCs resulting from the three  $\sigma_{\text{C-H}}$  orbitals. The inclusion of a peptide bond, leading to  $2^\bullet$ , modifies the nature of electronic states, because of (i) the formation of a hydrogen bond between the amide CO and the  $\text{NH}_3$  groups; (ii) the lowering of the point group symmetry from  $C_3$  to  $C_s$ ; and (iii) the emergence of low-energy vacant molecular orbitals (antisymmetric amide  $\pi^*$  and symmetric amide  $\sigma_{\text{N-H}}$ ). The intramolecular hydrogen bond induces an energy increase of the associated  $\sigma_{\text{N-H}}$  orbital. Therefore, in  $\text{GS}^2$ , the single electron is located only on the two free  $\sigma_{\text{N-H}}$  orbitals ( $\text{MO}_4$ ). Their antibonding combination ( $\text{MO}_5$ ), observed for  $\text{ES}^1$ , corresponds to  $\text{ES}^2$  in  $2^\bullet$ . We underline that the contribution of the amide  $\pi^*$  orbital is very low in  $\text{ES}^2$ .  $\text{ES}^3$ , which is located slightly below  $\text{ES}^2$  (0.85 eV vs 1.01 eV), can be described as the in-phase combination between  $\sigma_{\text{C-H}}$  ( $\text{ES}^3$ ) and the amide  $\sigma_{\text{N-H}}$  orbitals. This analysis suggests that, contrary to the traditional point of view in ECD and ETD,<sup>38</sup> the electron is not trapped into a Rydberg orbital on positive moieties of protonated amines but is rather located on their antibonding  $\sigma_{\text{N-H}}$  orbitals, and that the amide  $\pi^*$ -orbital is less accessible than an appropriate combination of the  $\sigma^*$ -orbitals. EOM-CCSD calculations (Figures 2 and 3) give very similar transition energies for both  $1^\bullet$  and  $2^\bullet$ , confirming the relevance of the CASSCF-MRCI data.

We have assessed the accuracies reached with several DFT approaches for  $1^\bullet$  and  $2^\bullet$ . The selected criteria were the absolute transition energies (from the GS), and the splitting between the excited states, considering the CASSCF-MRCI results as reference. The nature of all ESs has been established by comparison of the CASSCF-MRCI SOMO and the TD-DFT spin density distribution (see Figures 2 and 3), and reasonable match was obtained in the majority of cases. The TD-DFT vertical energies are compared to their CASSCF-MRCI counter-





**Figure 6.** (a) Spin densities of GS<sup>2</sup> and ES<sup>2,2</sup> for DFT methods. (b) CCSD SOMO of GS<sup>2</sup> and vacant MO with the largest contribution (0.887) in ES<sup>2,2</sup> at the EOM-CCSD level. (c) CASSCF-MRCI SOMO of GS<sup>2</sup> and ES<sup>2,2</sup>. Atomic spin densities of NH<sub>3</sub>, CH<sub>2</sub>, and CONHCH<sub>3</sub> are given in bold, italic, and underlined fonts, respectively.



**Figure 7.** Effects of the attenuation parameter  $\omega$  on the vertical excitation energy to ES<sup>2,2</sup> and its spin density located on the CONHCH<sub>3</sub> amide group (values in electron) using tuned LC-BLYP.

**Table 2.** States of Reduced 3<sub>10</sub><sup>++</sup> Obtained with  $\omega$ B97X-D and LC-BLYP Functionals and *Ab Initio* Methods

state	electronic structure similar to	Vertical Excitation Energies (eV)			
		LC-BLYP	$\omega$ B97X-D	EOM-CCSD <sup>a</sup>	CASSCF-MRCI
GS <sup>3d(a=10)</sup>	GS <sup>1</sup>	0.00	0.00	0.00	0.00
ES <sup>3d(a=10)1</sup>	GS <sup>2</sup>	0.56	0.55	0.56	0.57
ES <sup>3d(a=10)2</sup>	ES <sup>1,2</sup>	1.43	1.46	1.53	1.51
ES <sup>3d(a=10)3</sup>	ES <sup>1,1</sup>	1.45	1.47	1.54	1.52
ES <sup>3d(a=10)4</sup>	ES <sup>2,2</sup>	1.58	1.33	1.58	1.61

<sup>a</sup>The exact nature of the ESs could not be checked, as several vacant MOs contribute to their description.

parts in Figures 4 and 5. Qualitatively, for 1<sup>+</sup>, all RSHs and methods including 100%  $X^{\text{HF}}$  reproduce the ES ordering obtained with CASSCF-MRCI, but for M11 and M06-HF, which interchange two states (ES<sup>1,4</sup>, ES<sup>1,5</sup>) with ES<sup>1,6</sup>. All conventional functionals, except M06-2X, interchange (ES<sup>1,1</sup>, ES<sup>1,2</sup>) with ES<sup>1,3</sup>, but this effect is probably not very significant, because of the energetic proximity of these states at the CASSCF-MRCI level (see above). M06 also inverts ES<sup>1,4</sup> and ES<sup>1,5</sup> with ES<sup>1,6</sup>. For 2<sup>+</sup>, all functionals (but HISSbPBE and BH&HLYP) invert the first and second ESs. The mean absolute deviations (MAD) of the TD-DFT vertical excitation energies, MAD(E), as well as of the relative state ES energy deviations, MAD( $\Delta E$ ), are both reported in Table 1. As mentioned above, EOM-CCSD data are very similar to CASSCF-MRCI ones, leading to very small

MAD values between these methods. The largest MAD(E) values are observed for M06 and M11 for 1<sup>+</sup> and 2<sup>+</sup>, respectively, whereas most functionals yield mean deviations smaller than 0.1 eV, the most accurate being CAM-B3LYP, BH&HLYP, LC-BLYP, HFLYP,  $\omega$ B97,  $\omega$ B97X,  $\omega$ B97X-D, and M06-2X, as well as the Hartree–Fock (HF) method. BLYP, B3LYP, HISSbPBE, and LC-BPW91 functionals have a MAD(E) close to 0.1 eV for 1<sup>+</sup> but this value is significantly larger for 2<sup>+</sup>. The trends for the MAD( $\Delta E$ ) of 1<sup>+</sup> are similar to their MAD(E) counterparts. In contrast, for 2<sup>+</sup>, the trends differ for these two criteria. Indeed, the MAD( $\Delta E$ ) of BH&HLYP and HISSbPBE are very small, indicating a good match with CASSCF-MRCI, whereas most functionals have MAD( $\Delta E$ ) values close to 0.2 eV, except LC-BPW91, BLYP,  $\omega$ B97X-D, and  $\omega$ B97X, which yield larger deviations. By averaging these deviations for both 1<sup>+</sup> and 2<sup>+</sup>, we can conclude that (i) the most accurate methods, in terms of energies, are BH&HLYP, CAM-B3LYP, LC-BLYP, HFLYP, and  $\omega$ B97, and (ii) those providing the less-satisfying results are BLYP, M06, M11, and LC-BPW91.

The LC-BLYP spin densities are represented in Figures 2 and 3. For all ground and excited states, the shape of the SOMO and the spin density are very similar. The same qualitative agreement is obtained for all other methods used, except for, on the one hand, HF (see ES<sup>1,5</sup> in Figure S1 in the SI) and, on the other hand, M06-HF, M11, M06-2X, and M06, which give large spin contamination for ESs of both 1<sup>+</sup> and 2<sup>+</sup>, inducing unreasonable spin density distributions. Significant quantitative differences between the electronic states are observed for system 2<sup>+</sup>, as the spin density of many states is more delocalized with DFT, and more specifically with conventional functionals, than with CASSCF-MRCI SOMO or EOM-CCSD data. This is illustrated for both GS<sup>2</sup> and ES<sup>2,2</sup> in Figure 6. While the ground-state densities are qualitatively similar for all methods, quantitative estimation of the atomic spin density shows however that the smaller the  $X^{\text{HF}}$ , the larger the electron delocalization on the amide moiety. This effect is more pronounced for ES<sup>2,2</sup>. With conventional functionals, the spin density is delocalized on the entire molecule, mainly on the NH<sub>3</sub> and CO groups. An increase of  $X^{\text{HF}}$  induces a localization of the spin density on the CH<sub>2</sub>–NH<sub>3</sub> termination, as observed with CASSCF-MRCI and EOM-CCSD. The undue delocalization obtained with low  $X^{\text{HF}}$  functionals can be correlated with smaller excitation energies, thereby explaining the inaccuracy of these approaches from both qualitative and quantitative points of view (Figure 5).

With RSH, the same trends are observed. CAM-B3LYP, which does not include full  $X^{\text{HF}}$  at long range, shows a larger spin

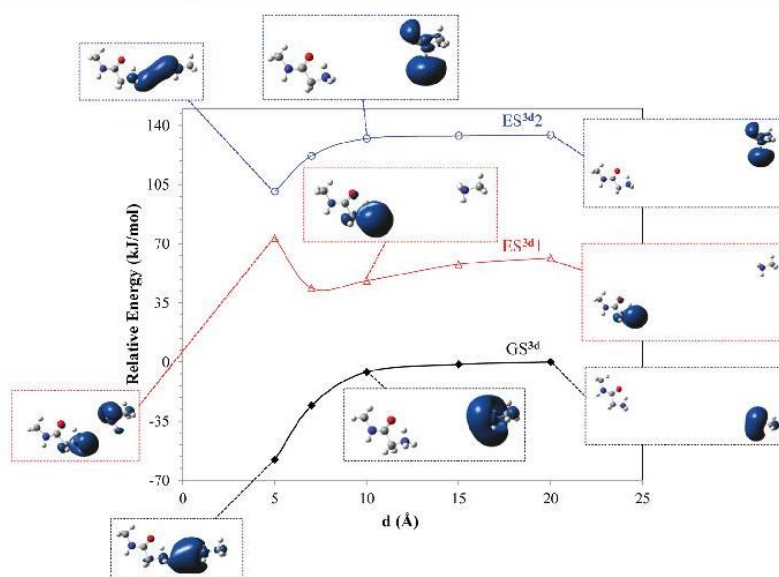


Figure 8. Effect of the N...N distance on the electronic structure of  $3_d^{4+}$  ( $d = 5, 7, 10, 15, 20$  Å). LC-BLYP spin densities and relative energies are given. Relative energies are calculated in comparison to the energy of  $GS^{3d(d=20)}$ .

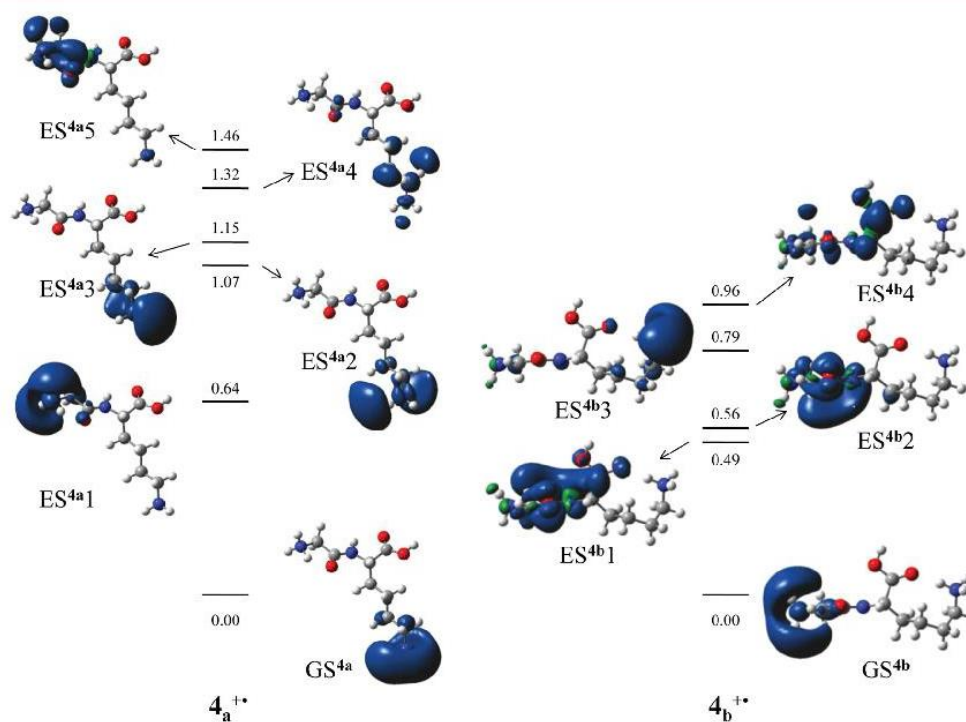


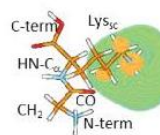
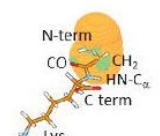
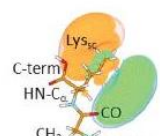

Figure 9. Spin densities and vertical excitation energies (eV) of ground and excited states of  $4_a^{4+}$  and  $4_b^{4+}$  at the LC-BLYP level.

delocalization on the amide group than LC-BLYP. The comparison of  $\omega$ B97,  $\omega$ B97X, and  $\omega$ B97X-D results (see Figure S2 in the SI) indicates a stronger delocalization for the last-mentioned, in accordance with its smaller attenuation parameter  $\omega$  (all three functionals presenting an asymptotically correct

$X^{\text{HF}}$ ). To ascertain this hypothesis, we have modified the parameter  $\omega$  of LC-BLYP and investigated the energy and spin density of ES<sup>2</sup> (see Figure 7). Similar calculations have been achieved for  $\omega$ B97X-D,  $\omega$ B97X, and  $\omega$ B97 (see Figure S3 in the SI). For  $\omega$  values between 0 and 0.3, the smaller the  $\omega$ , the



Table 3. Atomic Spin Densities of Ground and Excited States of  $4_a^{+\bullet}$  and  $4_b^{+\bullet}$ 

	N-term <sup>a</sup>	CH <sub>2</sub> <sup>b</sup>	CO <sup>c</sup>	HN-C <sub>α</sub> <sup>d</sup>	Lys <sub>SC</sub> <sup>e</sup>	C-term <sup>f</sup>	SOMO <sup>g</sup>
GS <sup>4a</sup>							
B3LYP	0.29	0.06	0.06	0.00	0.61	0.01	
M06-2X	0.04	0.01	0.01	0.00	0.94	0.00	
CAM-B3LYP	0.13	0.03	0.01	0.00	0.83	0.00	
ωB97X-D	0.00	0.00	0.00	0.00	1.00	0.00	
ωB97X-D <sup>h</sup>	0.00	0.00	0.00	0.00	1.00	0.00	
ωB97	0.00	0.00	0.00	0.00	1.00	0.00	
LC-BLYP	0.00	0.00	0.00	0.00	1.00	0.00	
HFLYP <sup>i</sup>	0.00	0.00	0.00	0.00	1.00	0.00	
ES <sup>4a1</sup>							
B3LYP	0.36	0.06	-0.02	0.01	0.63	-0.01	
M06-2X	0.08	0.02	0.01	0.01	0.95	0.00	
CAM-B3LYP	0.12	0.02	-0.01	0.00	0.88	0.00	
ωB97X-D	0.68	0.17	0.10	0.02	0.06	0.01	
ωB97X-D <sup>h</sup>	0.73	0.16	0.04	0.01	0.07	0.00	
ωB97	0.74	0.16	0.03	0.01	0.08	0.00	
LC-BLYP	0.78	0.16	0.02	0.01	0.04	0.00	
HFLYP <sup>i</sup>	0.78	0.16	0.02	0.01	0.05	0.00	
ES <sup>4a2</sup>							
B3LYP	0.00	0.11	0.46	0.03	0.59	0.01	
M06-2X	-0.01	0.14	0.64	0.04	0.34	0.00	
CAM-B3LYP	-0.02	0.10	0.58	0.03	0.70	0.01	
ωB97X-D	0.16	0.14	0.62	0.04	0.05	0.01	
ωB97X-D <sup>h</sup>	0.02	0.01	0.00	0.00	0.98	0.00	
ωB97	0.02	0.01	0.00	0.00	0.98	0.00	
LC-BLYP	0.01	0.00	0.00	0.00	0.98	0.00	
HFLYP <sup>i</sup>	0.01	0.00	0.00	0.00	0.99	0.00	
GS <sup>4b</sup>							
B3LYP	0.32	0.12	0.07	0.06	0.20	0.23	
M06-2X	0.51	0.18	0.05	0.05	0.04	0.16	
CAM-B3LYP	0.38	0.20	0.07	0.08	0.05	0.21	
ωB97X-D	0.02	0.06	0.31	0.10	0.01	0.51	
ωB97X-D <sup>h</sup>	0.77	0.19	0.02	0.01	0.00	0.01	
ωB97	0.76	0.20	0.02	0.01	0.00	0.01	
LC-BLYP	0.74	0.22	0.02	0.02	0.01	0.01	
HFLYP <sup>i</sup>	0.00	0.00	0.01	0.04	0.01	0.93	

<sup>a</sup>NH<sub>3</sub> terminal group. <sup>b</sup>Glycine C<sub>α</sub>H<sub>2</sub> group. <sup>c</sup>Glycine CO group. <sup>d</sup>Lysine HN-C<sub>α</sub> group. <sup>e</sup>Lysine (CH<sub>2</sub>)<sub>4</sub>NH<sub>3</sub> side chain. <sup>f</sup>COOH terminal group. <sup>g</sup>CASSCF-MRCI SOMO. <sup>h</sup>ωB97X-D with a modified parameter ω = 0.30. <sup>i</sup>See footnote g in Table 1.

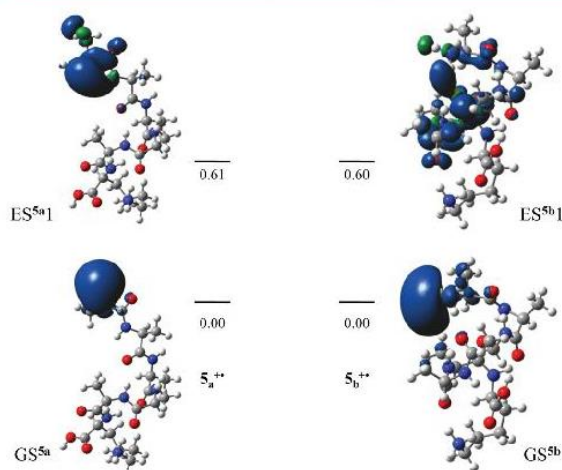
smaller the transition energies and the larger the spin density on the amide group. For  $\omega \geq 0.3$  ( $\omega \geq 0.4$ ), the transition energy (spin density) becomes nearly independent of the attenuation parameter.

Considering the results obtained for relative energies, MADs, and spin densities, we concluded that LC-BLYP, ωB97, and HFLYP produce the most satisfactory results. To study the larger systems  $3_d^{+\bullet}$ ,  $4_a^{+\bullet}$ ,  $4_b^{+\bullet}$ ,  $5_a^{+\bullet}$ , and  $5_b^{+\bullet}$ , we have used these methods as well as the commonly used B3LYP, M06-2X, ωB97X-D, and CAM-B3LYP hybrid functionals in order to confirm the trends obtained on model systems. The best match with CASSCF-MRCI and EOM-CCSD was obtained with LC-BLYP, which is therefore mainly used in the following (see the SI for additional results).

**Studies of Hydrogen-Rich Peptide Cation Radicals.** We first focus on  $3_d^{+\bullet}$ , which can be viewed as the combination of  $1^+$ ,  $2^+$ , and one additional electron.  $3_d^{+\bullet}$  is a simplified model of real hydrogen-rich peptide cation radicals. For example,  $3_{10}^{+\bullet}$  is an analogue of  $4_a^{+\bullet}$  as they present the same noncovalent interactions around the NH<sub>3</sub> groups and similar N...N distances (10.0 and 9.9 Å, respectively). Clearly, the electronic states of

$3_{10}^{+\bullet}$  can be viewed as the combination of those of  $1^+$  and  $2^+$  (see Table 2, Figure 8, and Figure S4 in the SI), the electron being localized on one of the two subgroups. Note that the fourth ES of  $3_{10}^{+\bullet}$  ( $= \text{ES}^{3d(d=10)}4$  in Table 2) is overstabilized with ωB97X-D, because the delocalization of the electron on the amide group is exaggerated, compared to LC-BLYP. This effect induces a reversal in the ES ordering. Similar incorrect results, indicating a too easily accessible amide  $\pi^*$ -orbital, are therefore expected for other peptides with methods including too small ω value and/or less than 100%  $X^{HF}$  at long range. Vertical excitation energies computed at the EOM-CCSD and CASSCF-MRCI levels give very good agreement with the LC-BLYP values (Table 2).

The variation of the distance between the ammonium groups allows us to study the influence of this chemical parameter on both the energy and spin density of each state. Results for GS<sup>3d</sup>, ES<sup>3d1</sup> and ES<sup>3d2</sup> are discussed here and those of ES<sup>3d3</sup> and ES<sup>3d4</sup> are reported in the SI. As expected, the energetic ranking and the nature of the ground and excited states obtained with LC-BLYP agree with the CASSCF-MRCI reference. Similar agreement is obtained with EOM-CCSD for the vertical excitation energies, even if larger values are observed in some cases, in particular for



**Figure 10.** LC-BLYP spin densities and vertical excitation energies of the GS and ES1 of  $S_a^{+*}$  and  $S_b^{+*}$ .

$3_s^{+*}$  (see the SI). The electronic structure of  $3_d^{+*}$  remains unchanged for  $d \geq 10$  Å, as shown in Figure 8. For the small distances, states of  $3_d^{+*}$  are obtained by symmetry-adapted linear combinations of the electronic states of  $1^+$  and  $2^+$ . In the ground state, the electron is delocalized between the two ammonium groups, corresponding to the in-phase combination of  $GS^1$  and  $GS^2$ , whereas  $ES^{3d}1$  corresponds to the out-of-phase combination. Similarly,  $ES^{3d}2$  is the in-phase mix of  $ES^11$  and  $ES^22$ . The nature of these states explains their stabilization (in-phase combination) or destabilization (out-of-phase combination), compared to the situation observed for  $d \geq 10$  Å.

We used LC-BLYP to investigate the electronic structure of  $4_a^{+*}$  and  $4_b^{+*}$  (Figure 9). As expected,  $4_a^{+*}$  behaves similarly to  $3_{10}^{+*}$ . The lowest electronic states present spin densities localized on one of the ammonium group. The first excited state in which the amide  $\pi^*$ -orbital is populated is  $ES^{4a}5$  at 1.46 eV, whereas  $ES^{4a}1$  at 0.64 eV shows very low spin density on the amide moiety. Once again,  $\omega$ B97X-D, CAM-B3LYP, B3LYP,<sup>42</sup> and M06-2X (see the SI) give a populated amide  $\pi^*$ -orbital for an ES of much lower energy (at 0.99, 0.61, 0.40, and 0.47 eV, respectively), in sharp disagreement with the CASSCF-MRCI reference, which give results similar to LC-BLYP. Comparison of the atomic spin densities for the GS and the first ESs of  $4_a^{+*}$  (Table 3) quantitatively confirm the performance of the functionals.

For  $4_b^{+*}$ , another low-energy conformation of the GlyLys dipeptide that includes one additional hydrogen bond between the COOH and the  $NH_3$  group of the lysine, the nature of the electronic states is completely modified. Indeed, in the ground state  $GS^{4b}$ , the electron is located on the N-terminal ammonium, similar to what is found with the first excited state ( $ES^{4a}1$ ) of  $4_a^{+*}$ . The state corresponding to  $GS^{4a}$  is pushed up to  $ES^{4b}3$  due to the additional hydrogen bond. The lowest excited state in which the electron is localized in the amide  $\pi^*$ -orbital appears 0.49 eV above the ground-state, and is therefore significantly more accessible than in  $4_a^{+*}$ . Among the other tested functionals (see Table 3 and the SI), only  $\omega$ B97 gives the same ground state as LC-BLYP and CASSCF-MRCI. CAM-B3LYP, B3LYP, and M06-2X lead to an extensive delocalization of the spin density over the N-term ammonium, the C-term carboxylic group, and the amide group of the dipeptide in  $GS^{4b}$ . The HFLYP ground

state presents the unpaired electron located in the carboxylic group rather than on the N-term  $NH_3$ . Eventually, because of the overstabilization of the electronic states including partly filled amide  $\pi^*$ -orbitals,  $\omega$ B97X-D inverts  $GS^{4b}$  and  $ES^{4b}1$ . This reversal can be solved by increasing the attenuation parameter from 0.2 a.u. to 0.3 a.u. (functional noted  $\omega$ B97X-D\*), which restores accurate results for both  $4_b^{+*}$  and  $4_a^{+*}$ .

Are the above conclusions on the nature of the electronic states and on the reliability of DFT functionals in that framework maintained for longer, more-realistic, peptides? To explore this issue, our study has been extended to the hexapeptide AlaLys, which has been examined recently in a joint theoretical-experimental study and was shown to undergo cleavages of the N-C $\alpha$  bond with ECD and ETD.<sup>15</sup> The spin density of the  $GS^{Sa/b}$  and the  $ES^{Sa/b}1$  for two low-energy conformations are given in Figure 10. In  $S_a^{+*}$ , the N-term ammonium is isolated from the other groups while the ammonium group of the lysine side chain is hydrogen-bonded to both the COOH and a CO group of one amide.  $S_b^{+*}$  also includes a hydrogen bond between the N-term ammonium and one CO. In both cases, at the LC-BLYP level (Figure 10), as well as for  $\omega$ B97 and HFLYP (see the SI), the ground state corresponds to the addition of the single electron to the N-term ammonium, the charged site with the larger electron affinity, that is, with the weaker hydrogen bonds. The spin density in  $ES^{Sb}1$  is located around the amide bond closest to the N-term, whereas a spin-density delocalization over several amide  $\pi^*$ -orbitals is predicted for  $ES^{Sb}1$ .  $\omega$ B97X-D, CAM-B3LYP, and B3LYP give similar results for  $S_a^{+*}$  with, however, larger spin density delocalization, and smaller transition energy in particular for B3LYP. As for  $4_b^{+*}$ , ground and excited states of  $S_b^{+*}$  are not correctly described by these methods as well as by M06-2X, confirming all the trends noted previously.

## CONCLUSIONS

Using CASSCF-MRCI results as benchmarks, as well as EOM-CCSD calculations, we have benchmarked a panel of exchange-correlation functionals in the framework of the description of the ground and excited states of hydrogen-rich peptide cation radicals. Fifteen functionals have been assessed considering two criteria on small chemical models: (i) energetics, including the ordering of excited states, vertical excitation energies and the excited states splitting; and (ii) spin densities of both ground and excited states. All conventional DFT approaches fail to pass these two tests, whereas LC-BLYP,  $\omega$ B97, and HFLYP provide the best agreement with CASSCF-MRCI and EOM-CCSD. Notably, in range-separated hybrids, the attenuation parameter  $\omega$  can significantly modify both the spin density distribution and the vertical excitation energies even when 100%  $X^{HF}$  is included at long-range interelectronic distance.

Seven functionals have been applied on hydrogen-rich di- and hexa-peptide cation radicals, confirming both the robustness of  $\omega$ B97 and LC-BLYP, as well as the failure of the popular B3LYP, M06-2X,  $\omega$ B97X-D approaches, and, to a smaller extent, CAM-B3LYP. Increasing the size of the studied systems as well as changing their conformation allowed us to understand the nature and the order of the electronic states as a function of noncovalent interactions around the ammonium groups. It turned out that the number of hydrogen bonds on the ammonium groups is a decisive chemical parameter that guides the electronic structure of hydrogen-rich peptide radicals.



## ■ ASSOCIATED CONTENT

## ■ Supporting Information

Complete refs S9 and S6. Assessment of the active space size for CASSCF-MRCI calculations and influence of the basis set on EOM-CCSD results. HF electronic structure of **1**<sup>+</sup>. Effects of  $X^{HF}$  and attenuation parameter on the GS<sup>2</sup> and ES<sup>2</sup> properties using  $\omega$ B97,  $\omega$ B97X, and  $\omega$ B97X-D. LC-BLYP relative energies and spin densities for ES<sup>3d</sup>3 and ES<sup>3d</sup>4. Relative energies and SOMOs of  $3d^{+*}$ ,  $4a^{+*}$ , and  $4b^{+*}$  obtained at the CASSCF-MRCI and EOM-CCSD levels. The  $\omega$ B97X-D\* ( $\omega = 0.3$ ) electronic structure of  $4a^{+*}$  and  $4b^{+*}$ . Spin densities and vertical excitation energies of  $4a^{+*}$ ,  $4b^{+*}$ ,  $5a^{+*}$ , and  $5b^{+*}$  obtained with  $\omega$ B97X-D,  $\omega$ B97, HFLYP, CAM-B3LYP, B3LYP, and M06-2X. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

## Corresponding Author

\*Tel.: +33 (0)1 69 33 48 34. Fax: +33 (0)1 69 33 48 03. E-mail: [gilles.frison@polytechnique.org](mailto:gilles.frison@polytechnique.org).

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

V.R., D.J., and G.F. gratefully acknowledge financial support from GDR 3533 "EMIE". D.J. acknowledges the European Research Council (ERC) and the Région des Pays de la Loire for financial support in the framework of a Starting Grant (Marches-278845) and a *recrutement sur poste stratégique*, respectively. This work was performed using HPC resources from GENCI-CINES (Grant No. x2013087105).

## ■ REFERENCES

- (1) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
- (2) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245–249.
- (3) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
- (4) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563–573.
- (5) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24*, 201–222.
- (6) Mikesch, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta* **2006**, *1764*, 1811–1822.
- (7) Zubarev, R. A.; Haselmann, K. F.; Budnik, B.; Kjeldsen, F.; Jensen, F. *Eur. J. Mass Spectrom.* **2002**, *8*, 337–349.
- (8) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1793–1800.
- (9) Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T. *Anal. Chem.* **1999**, *71*, 4250–4253.
- (10) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Anal. Chem.* **1999**, *71*, 4431–4436.
- (11) Shi, S. D.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73*, 19–22.
- (12) Sze, S. K.; Ge, Y.; Oh, H. B.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774–1779.
- (13) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672–678.
- (14) Turecek, F.; Julian, R. R. *Chem. Rev.* **2013**, *113*, 6691–6733.
- (15) Wodrich, M.; Zhurov, K.; Vorobyev, A.; Ben Hamidane, H.; Corminboeuf, C.; Tsybin, Y. J. *Phys. Chem. B* **2012**, *116*, 10807–10815.

- (16) Wodrich, M. D.; Zhurov, K. O.; Corminboeuf, C.; Tsybin, Y. O. *J. Phys. Chem. B* **2014**, *118*, 2985–2992.
- (17) Zhurov, K. O.; Wodrich, M. D.; Corminboeuf, C.; Tsybin, Y. O. *J. Phys. Chem. B* **2014**, *118*, 2628–2637.
- (18) Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A. *Int. J. Mass Spectrom.* **2006**, *248*, 124–135.
- (19) Syrtstad, E. A.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 208–224.
- (20) Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J. *J. Phys. Chem. B* **2003**, *107*, 13505–13511.
- (21) Sobczyk, M.; Anusiewicz, I.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J. *J. Phys. Chem. A* **2005**, *109*, 250–258.
- (22) Turecek, F.; Chung, T. W.; Moss, C. L.; Wyer, J. A.; Ehlerding, A.; Holm, A. I. S.; Zettergren, H.; Nielsen, S. B.; Hvelplund, P.; Chamot-Rooke, J.; Bythell, B.; Paizs, B. *J. Am. Chem. Soc.* **2010**, *132*, 10728–10740.
- (23) Turecek, F.; Moss, C. L.; Chung, T. W. *Int. J. Mass Spectrom.* **2012**, *330–332*, 207–219.
- (24) Breuker, K.; Oh, H. B.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *Eur. J. Mass Spectrom.* **2002**, *8*, 177–180.
- (25) Cooper, H. J.; Hudgins, R. R.; Hakansson, K.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 241–249.
- (26) Cooper, H. J. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1932–1940.
- (27) Fung, Y. M.; Chan, T. W. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1523–1535.
- (28) Lee, S.; Chung, G.; Kim, J.; Oh, H. B. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3167–3175.
- (29) Liu, H.; Hakansson, K. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2007–2013.
- (30) Haselmann, K. F.; Schmidt, M. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1003–1008.
- (31) Falth, M.; Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Andren, P. E.; Zubarev, R. A. *Anal. Chem.* **2008**, *80*, 8089–8094.
- (32) van der Rest, G.; Hui, R.; Frison, G.; Chamot-Rooke, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1631–1644.
- (33) Marek, A.; Pepin, R.; Peng, B.; Laszlo, K. J.; Bush, M. F.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1641–1653.
- (34) Kalopothakis, J. M. D.; Berezovskaya, Y.; Zampronio, C. G.; Faull, P. A.; Barran, P. E.; Cooper, H. J. *Chem. Commun.* **2014**, *50*, 198–200.
- (35) Alizadeh, E.; Sanche, L. *Chem. Rev.* **2012**, *112*, 5578–5602.
- (36) Gu, J.; Leszczynski, J.; Schaefer, H. F. *Chem. Rev.* **2012**, *112*, 5603–5640.
- (37) Neff, D.; Simons, J. *J. Phys. Chem. A* **2009**, *114*, 1309–1323.
- (38) Simons, J. *Chem. Phys. Lett.* **2010**, *484*, 81–95.
- (39) Simons, J. *J. Am. Chem. Soc.* **2010**, *132*, 7074–7085.
- (40) Chamot-Rooke, J.; Malosse, C.; Frison, G.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2146–2161.
- (41) Frison, G.; van der Rest, G.; Turecek, F.; Besson, T.; Lemaire, J.; Maitre, P.; Chamot-Rooke, J. *J. Am. Chem. Soc.* **2008**, *130*, 14916–14917.
- (42) Turecek, F.; Chen, X.; Hao, C. *J. Am. Chem. Soc.* **2008**, *130*, 8818–8833.
- (43) Panja, S.; Brøndsted Nielson, S.; Hvelplund, P.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1726–1742.
- (44) Jensen, C. S.; Holm, A. I. S.; Zettergren, H.; Overgaard, J. B.; Hvelplund, P.; Brøndsted Nielsen, S. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1881–1889.
- (45) Sohn, C. H.; Chung, C. K.; Yin, S.; Ramachandran, P.; Loo, J. A.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2009**, *131*, 5444–5459.
- (46) Bythell, B. J. *J. Phys. Chem. A* **2013**, *117*, 1189–1196.
- (47) Maclot, S.; Rangama, J.; Brøndsted Nielson, S.; Pouilly, J. C. *Int. J. Mass Spectrom.* **2013**, *337*, 1–11.
- (48) Pepin, R.; Laszlo, K. J.; Peng, B.; Marek, A.; Bush, M. F.; Turecek, F. *J. Chem. Phys. A* **2014**, *118*, 308–324.
- (49) Moss, C. L.; Liang, W.; Li, X.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 446–459.
- (50) Ledvina, A. R.; Chung, T. W.; Hui, R.; Coon, J. J.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1351–1363.



- (51) Gilson, A. I.; van der Rest, G.; Chamot-Rooke, J.; Kurlancheek, W.; Head-Gordon, M.; Jacquemin, D.; Frison, G. *J. Phys. Chem. Lett.* **2011**, *2*, 1426–1431.
- (52) (a) Lundberg, M.; Siegbahn, P. E. M. *J. Chem. Phys.* **2005**, *122*, 224103–224109. (b) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. *Science* **2008**, *321*, 792–794.
- (53) Johnson, E. J.; Salamone, M.; Bietti, M.; DiLabio, G. A. *J. Phys. Chem. A* **2013**, *117*, 947–952.
- (54) (a) Conradie, J.; Ghosh, A. J. *J. Phys. Chem. B* **2007**, *111*, 12621–12624. (b) Boguslawski, K.; Jacob, C. R.; Reiher, M. *J. Chem. Theory Comput.* **2011**, *7*, 2740–2752.
- (55) (a) Leang, S. S.; Zahariev, F.; Gordon, M. S. *J. Chem. Phys.* **2012**, *136*, 104101–104112. (b) Laurent, A. D.; Jacquemin, D. *Int. J. Quantum Chem.* **2013**, *113*, 2019–2039 and references therein.
- (56) (a) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110. (b) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *129*, 104103.
- (57) Goerigk, L.; Grimme, S. *J. Chem. Phys.* **2010**, *132*, 184103.
- (58) (a) Vlcek, A.; Zalis, S. *Coord. Chem. Rev.* **2007**, *251*, 258–287. (b) Gonzalez, L.; Escudero, D.; Serrano-Andrés, L. *ChemPhysChem* **2012**, *13*, 28–51.
- (59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2010.
- (60) Yanai, T.; Tew, D.; Handy, N. *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (61) (a) Henderson, T. M.; Izmaylov, A. F.; Scuseria, G. E.; Savin, A. *J. Chem. Phys.* **2007**, *127*, 221103. (b) Henderson, T. M.; Izmaylov, A. F.; Scuseria, G. E.; Savin, A. *J. Chem. Theory Comput.* **2008**, *4*, 1254–1262.
- (62) Chai, J. D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (63) Chai, J. D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106–084115.
- (64) Likura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- (65) Tsuneda, T.; Hirao, K. *WIREs Comput. Mol. Sci.* DOI: 10.1002/wcms.1178.
- (66) Peverati, R.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.
- (67) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 5121–5129.
- (68) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126–13130.
- (69) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (70) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200–206.
- (71) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–100.
- (72) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–77.
- (73) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (74) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (75) Rabilloud, F. *J. Phys. Chem. A* **2013**, *117*, 4267–4278.
- (76) (a) Refaely-Abramson, S.; Baer, R.; Kronik, L. *Phys. Rev. B* **2011**, *84*, 0751441–8. (b) Kronik, L.; Stein, T.; Refaely-Abramson, S.; Baer, R. *J. Chem. Theory Comput.* **2012**, *8*, 1515–1531. (c) Refaely-Abramson, S.; Sharifzadeh, S.; Govind, N.; Autschbach, J.; Neaton, J. B.; Baer, R.; Kronik, L. *Phys. Rev. Lett.* **2012**, *109*, 226405 1–6.
- (77) Karolewski, A.; Kronik, L.; Kümmel, S. *J. Chem. Phys.* **2013**, *138*, 204115 1–11.
- (78) Agrawal, P.; Tkatchenko, A.; Kronik, L. *J. Chem. Theory Comput.* **2013**, *9*, 3473–3478.
- (79) Moore, B., II; Autschbach, J. *J. Chem. Theory Comput.* **2013**, *9*, 4991–5003.
- (80) Sun, H.; Autschbach, J. *ChemPhysChem* **2013**, *14*, 2450–2461.
- (81) Baer, R.; Livshits, E.; Salzner, U. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.
- (82) Jacquemin, D.; Moore, B., II; Planchat, A.; Adamo, C.; Autschbach, J. *J. Chem. Theory Comput.* **2014**, *10*, 1677–1685.
- (83) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (84) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157–173.
- (85) (a) Knowles, P. J.; Werner, H. J. *Chem. Phys. Lett.* **1988**, *145*, 514–522. (b) Werner, H. J.; Knowles, P. J. *J. Chem. Phys.* **1988**, *89*, 5803–5814.
- (86) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; et al. *Molpro, version 2010.1*, a package of ab initio programs, 2010. See <http://www.molpro.net>.
- (87) Langhoff, S. R.; Davidson, E. R. *Int. J. Quantum Chem.* **1974**, *8*, 61–72.
- (88) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

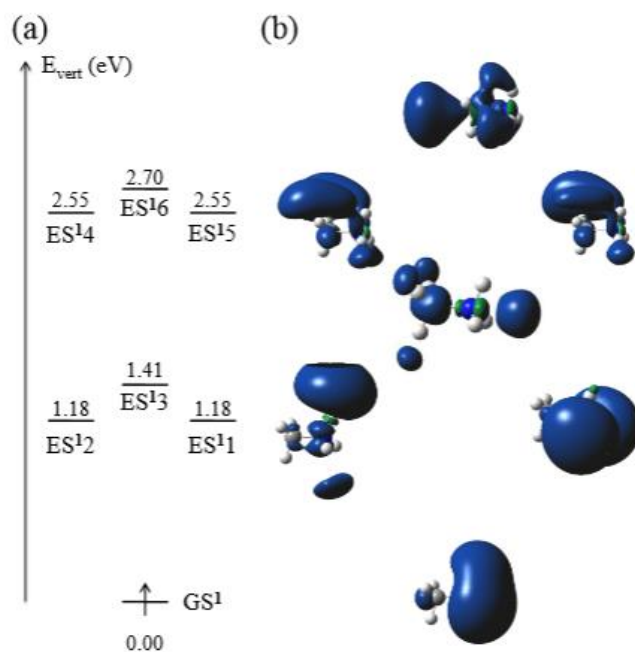
**Table S1.** Accuracy of CASSCF-MRCI calculations.<sup>a</sup>

<b>1<sup>*</sup></b>	(1,6) <sub>6</sub>	(1,10) <sub>7</sub>		
E(ES <sup>1</sup> 1)	1.46	1.46		
E(ES <sup>1</sup> 2)	1.46	1.46		
E(ES <sup>1</sup> 3)	1.47	1.47		
E(ES <sup>1</sup> 4)	2.67	2.67		
E(ES <sup>1</sup> 5)	2.67	2.67		
<b>2<sup>*</sup></b>	(1,3) <sub>3</sub>	(1,10) <sub>7</sub>		
E(ES <sup>2</sup> 1)	0.86	0.85		
E(ES <sup>2</sup> 2)	0.97	1.01		
<b>3<sub>5</sub><sup>++</sup></b>	(1,3) <sub>3</sub>	(1,5) <sub>3</sub>	(1,7) <sub>3</sub>	(5,5) <sub>3</sub>
E(ES <sup>3d(d=5)</sup> 1)	1.41	1.41	1.41	1.41
E(ES <sup>3d(d=5)</sup> 2)	1.70	1.70	1.70	1.70
<b>3<sub>10</sub><sup>++</sup></b>	(1,3) <sub>3</sub>	(1,5) <sub>5</sub>	(1,7) <sub>7</sub>	
E(ES <sup>3d(d=10)</sup> 1)	0.57	0.57	0.57	
E(ES <sup>3d(d=10)</sup> 2)	1.51	1.51	1.51	

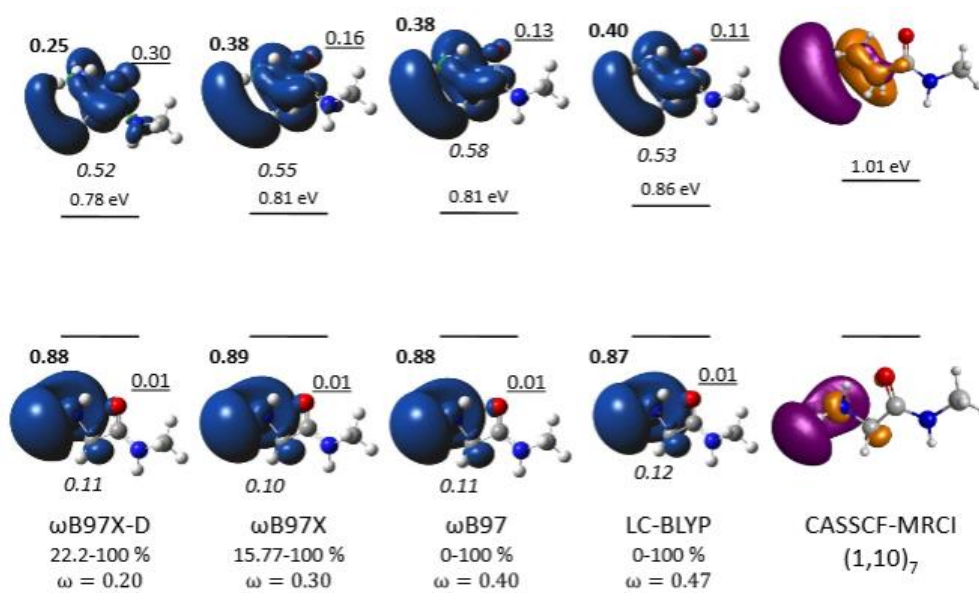
<sup>a</sup> Vertical excitation energies in eV**Table S2.** Influence of the basis set on EOM-CCSD results.<sup>a</sup>

	ES1	ES2	ES3	ES4	ES5	ES6
<b>1<sup>*</sup></b>						
6-311++G(2d,p)	1.48	1.48	1.48	2.68	2.68	2.82
aug-cc-pVTZ	1.41	1.41	1.41	2.52	2.52	2.53
<b>2<sup>*</sup></b>						
6-311++G(2d,p)	0.86	0.96	1.56	1.82	1.96	2.24
aug-cc-pVTZ	0.84	0.91	1.43	1.69	1.79	2.08

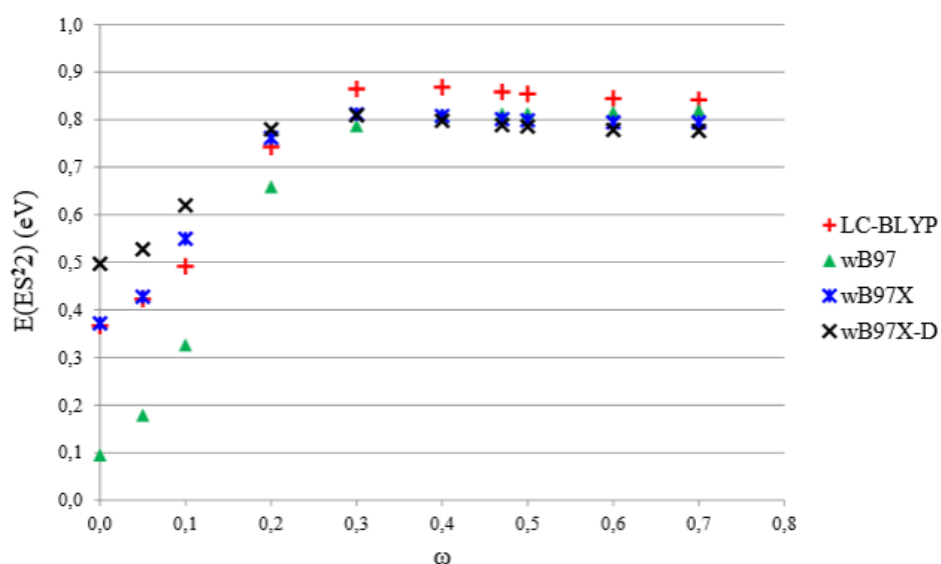
<sup>a</sup> Vertical excitation energies in eV



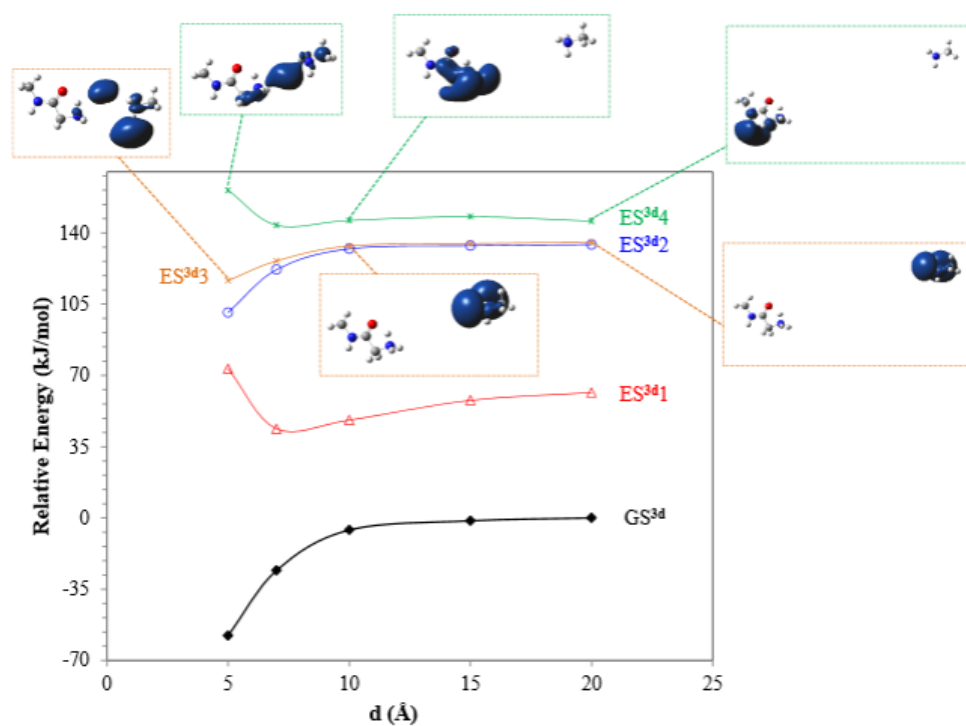
**Figure S1.** Vertical electronic states of **1<sup>\*</sup>**. (a) HF energies; (b) HF and TD-HF spin density distributions



**Figure S2.** Effect of  $X^{\text{HF}}$  on the spin densities of GS<sup>2</sup> and ES<sup>2</sup>. Atomic spin densities of NH<sub>3</sub>, CH<sub>2</sub> and CONHCH<sub>3</sub> are given in bold, italic and underlined fonts respectively.



**Figure S3.** Effects of the attenuation parameter  $\omega$  on the vertical excitation energy to  $\text{ES}^2$  using LC-BLYP,  $\omega$ B97,  $\omega$ B97X and  $\omega$ B97X-D.



**Figure S4.** Effect of the N...N distance on the electronic structure of  $3_d^{+\bullet}$  ( $d = 5, 7, 10, 15, 20$  Å). LC-BLYP spin densities and relative energies are given. Relative energies are calculated using the energy of  $\text{GS}^{3d(d=20)}$  as reference.

## IV.4. Les liaisons hydrogène, un facteur clé de la structure électronique des peptides radicaux cations

Nous avons souligné ci-dessus la disparité des résultats obtenus avec différentes fonctionnelles tant pour les états fondamentaux qu'excités des peptides radicaux cations. En confrontant les résultats DFT et TD-DFT avec les niveaux de références CASSCF-MRCI et EOM-CCSD, nous avons souligné que le paramètre d'atténuation a un effet significatif sur les densités de spin et sur les énergies relatives des états excités. Pour un paramètre d'atténuation trop faible, la délocalisation électronique sur les orbitales  $\pi_{CO}^*$  est artificiellement favorisée même si 100 % d'échange exact est appliqué à longue distance inter-électronique. De ce fait, nous avons conclu que LC-BLYP et  $\omega$ B97 sont deux fonctionnelles de choix pour l'étude de nos systèmes et elles seront utilisées par la suite. Etant donné qu'elles donnent des résultats similaires, nous avons discuté uniquement des résultats LC-BLYP.

La comparaison de la structure électronique de  $4_a^{+\bullet}$  et  $4_b^{+\bullet}$  puis  $5_a^{+\bullet}$  et  $5_b^{+\bullet}$  prouve que le site de capture de l'électron ne dépend pas seulement de la nature du peptide considéré, mais aussi de sa conformation, et particulièrement du réseau de liaisons hydrogène intramoléculaires entourant les sites chargés. Dans ce qui suit, nous visons à comprendre l'influence de la présence d'une ou plusieurs liaison(s) hydrogène sur la structure électronique de ces intermédiaires en utilisant différents environnements locaux autour des ammoniums (voir Figure IV.9).

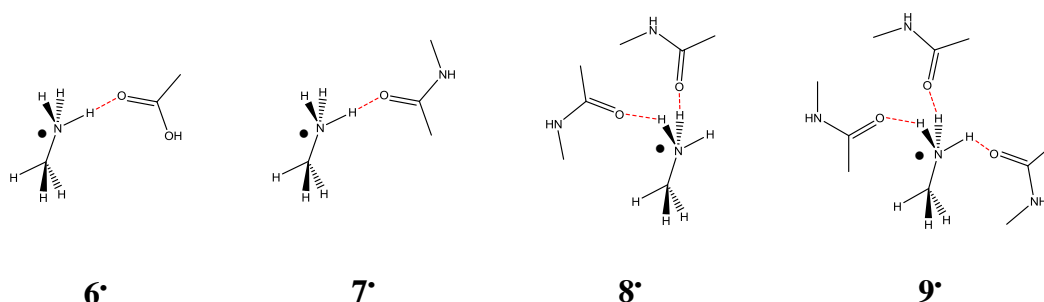


Figure IV.9. Nouveaux systèmes modèles étudiés. Les liaisons hydrogène intermoléculaires sont représentées en pointillés rouges.

### IV.4.1. Etudes de systèmes modèles.

La Figure IV.10 donne les représentations de la densité de spin pour  $GS^1$ ,  $ES^1$ ,  $ES^2$ ,  $GS^2$  et  $ES^2$  obtenues en LC-BLYP dans la partie IV.2. Ces densités de spin sont principalement localisées sur les  $\sigma_{N-H}^*$ . L'interaction de la  $\sigma_{N-H}^*$  de  $1^\bullet$  avec la paire libre d'un groupe carbonyle conduit à une diminution du nombre de  $\sigma_{N-H}^*$  disponible dans  $2^\bullet$ . Les combinaisons anti-liantes associées à  $ES^1$  et  $ES^2$  sont équivalentes et indiquent qu' $ES^2$  est peu influencé par la présence du carbonyle. Ainsi, l'énergie de recombinaison correspondant à la différence en énergie entre l'état fondamentale de  $1^+$  ( $GS^{1+}$ ) et  $ES^1$  est similaire à celle correspondant entre l'état fondamental de  $2^+$  ( $GS^{2+}$ ) et  $ES^2$  (2.61 vs 2.53 eV, respectivement). On constate que l'électron se répartit dans les trois  $\sigma_{N-H}^*$  pour le système  $1^\bullet$  et dans deux pour le système  $2^\bullet$ , ce qui correspond aux  $\sigma_{N-H}^*$  non solvatées. Cet effet se traduit par une diminution de l'énergie de recombinaison de  $1^\bullet$  vers  $2^\bullet$  (3.99 vs 3.39 eV). Nous avons voulu vérifier que cette observation est vraie lorsqu'on augmente le nombre de liaisons hydrogène environnant l'ammonium. Ainsi, plusieurs environnements locaux du groupement ammonium ont été reproduit par l'addition successive de N-méthyl-acétamide (noté NMA), qui est un petit système possédant une liaison peptidique, et ce jusqu'à complète solvatation du groupement  $NH_3$  (voir Tableau IV.1).

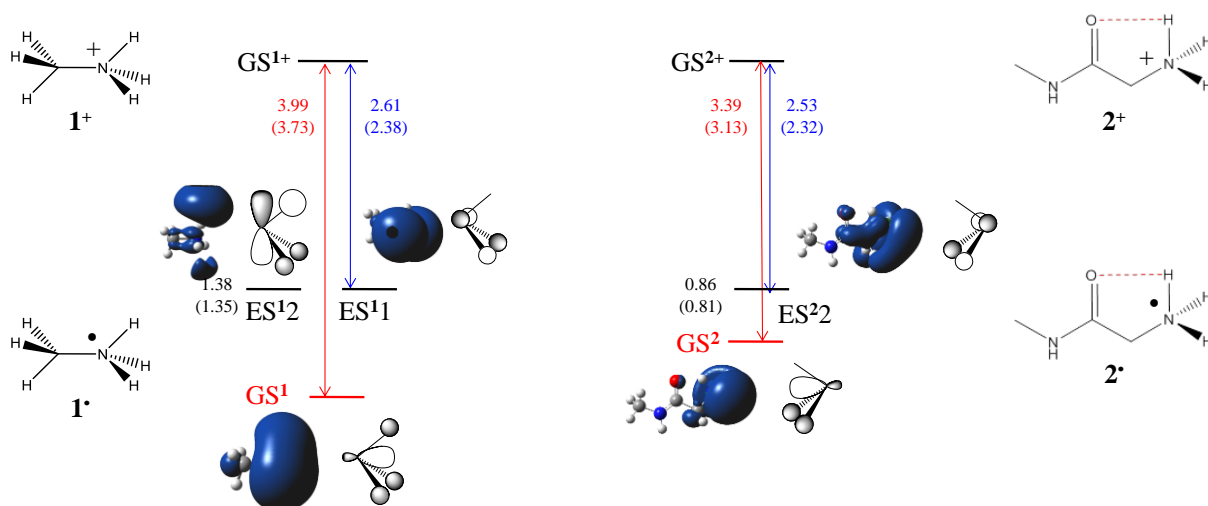

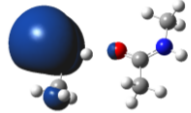

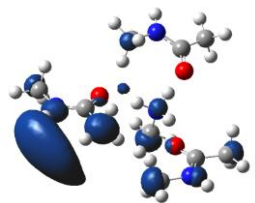


Figure IV.10. Effet d'une liaison hydrogène sur les densités de spin. Les énergies verticales (valeurs en noir) et les ERs (valeurs en rouge et bleu) sont calculées au niveau LC-BLYP et sont exprimées en eV.

Tableau IV.1. Evolution des  $ER_{\text{vert}}$  en fonction du nombre de  $\sigma_{\text{N-H}}^*$  disponible.

systèmes	$\sigma_{\text{N-H}}^*$ accessible	densité de spin de l'EF	$ER_{\text{vert}}$ (eV)
<b>1<sup>•</sup></b>	3		3.99 (3.73)
<b>7<sup>•</sup></b>	2		2.81 (2.58)
<b>8<sup>•</sup></b>	1		2.06 (1.84)
<b>9<sup>•</sup></b>	0		1.64 (1.40)

Comme dans **2<sup>•</sup>**, l'électron se localise sur les  $\sigma_{\text{N-H}}^*$  disponibles pour les systèmes **1<sup>•</sup>**, **7<sup>•</sup>** et **8<sup>•</sup>**, comme le montre les figures de densité de spin représentées à la Table IV.1. L'ammonium du système **9<sup>•</sup>** est totalement solvato (toutes les  $\sigma_{\text{N-H}}^*$  interagissent avec une paire libre d'un oxygène d'un carbonyle) et l'électron va se localiser sur un autre groupement. C'est ce qu'illustre la Figure IV.11 où la densité de spin de l'état fondamental de **9<sup>•</sup>** est localisée sur le NMA. Cette figure de densité de spin est similaire à celle du NMA seul dans son état fondamental (Figure IV.11).

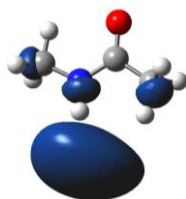


Figure IV.11. Densité de spin de l'état fondamental du NMA calculée au niveau LC-BLYP.

L'ER calculée au niveau LC-BLYP diminue avec l'augmentation du nombre de liaisons hydrogène formées autour de l'ammonium. Dans un composé présentant plusieurs sites, l'ammonium le moins solvato devrait donc capter préférentiellement l'électron. Toutefois, l'écart des  $ER_{\text{vert}}$  entre les modèles diminue à chaque ajout d'une nouvelle liaison hydrogène

(Table IV.2), rendant ainsi la prédiction sur la localisation de la densité de spin dans un peptide plus difficile. Cet écart vaut 1.18, 0.75 et 0.42 eV entre **1**<sup>•</sup> et **7**<sup>•</sup>, **7**<sup>•</sup> et **8**<sup>•</sup> et **8**<sup>•</sup> et **9**<sup>•</sup>, respectivement. Cet effet est également observé lors du peuplement de l'orbitale  $\pi_{CO}^*$  par l'électron pour un peptide contenant à la fois des sites de type **2**<sup>•</sup> et **6**<sup>•</sup> (1.57 et 1.53 eV, respectivement) ou bien à la fois **7**<sup>•</sup>, **8**<sup>•</sup> et **9**<sup>•</sup> (0.60, 0.39 et 0.13 eV, respectivement). Cela a des implications pour le mécanisme d'Utah-Washington car le nombre de liaisons hydrogène autour des ammoniums a également une influence sur le site de la capture électronique lorsque plusieurs  $\pi_{CO}^*$  existent dans l'intermédiaire radical cation.

Table IV.2. Energies de recombinaison pour des systèmes modèles dans le cas où les états  $GS^X$  et  $ES^X(\pi_{CO}^*)$  capturent l'électron.

Systèmes ( <b>X</b> <sup>•</sup> )	LC-BLYP		$\omega$ B97	
	$GS^X$	$ES^X(\pi_{CO}^*)$	$GS^X$	$ES^X(\pi_{CO}^*)$
<b>2</b> <sup>•</sup>	3.39	1.57	3.13	1.46
<b>6</b> <sup>•</sup>	3.13	1.53	2.88	1.37
<b>7</b> <sup>•</sup>	2.81	0.60	2.58	0.45
<b>8</b> <sup>•</sup>	2.06	0.39	1.84	0.20
<b>9</b> <sup>•</sup>	1.64	0.13	1.40	0.02
<b>3<sub>d</sub></b> <sup>++</sup>				
<b>3<sub>5</sub></b> <sup>++</sup>	6.90	3.61	6.64	3.42
<b>3<sub>7</sub></b> <sup>++</sup>	6.00	3.28	5.77	3.15
<b>3<sub>10</sub></b> <sup>++</sup>	5.33	2.90	5.08	2.78
<b>3<sub>15</sub></b> <sup>++</sup>	4.90	2.53	4.64	2.36
<b>3<sub>20</sub></b> <sup>++</sup>	4.68	2.32	4.42	2.16

La distance entre deux ammoniums (interaction coulombienne répulsive) au sein d'un peptide est un autre paramètre pouvant influencer la capture de l'électron. En effet, leur proximité augmente la valeur de l' $ER_{vert}$  du système comme le montre l'évolution de cette grandeur au sein des systèmes **3<sub>d</sub>**<sup>++</sup> (Table IV.2). Ainsi, l'environnement local d'un ammonium est perturbé par la proximité d'un autre ammonium, et cela rend plus difficile la prédiction du site de localisation de l'électron pour des environnements locaux distincts d'ammonium.

## IV.4.2 Applications

Nous avons comparé les  $ER_{vert}$  des ammoniums séparés pour les systèmes **4<sub>a</sub>**<sup>++</sup> et **4<sub>b</sub>**<sup>++</sup> (voir Figure IV.12). Le groupe  $NH_3$  de la lysine de **4<sub>a</sub>**<sup>++</sup> possède la plus forte  $ER_{vert}$ , donc il est attendu que ce groupement capture l'électron lorsque **4<sub>a</sub>**<sup>++</sup> est dans son état électronique



fondamental. Cette conclusion est confirmée par les densités de spin de l'état fondamental et des états excités des systèmes  $4_a^{++}$  et  $4_b^{++}$  calculées dans la partie IV.2. Similairement pour  $4_b^{++}$ , l'électron est également localisé sur le N-term  $\text{NH}_3$  qui a la plus forte  $\text{ER}_{\text{vert}}$  (3.39 eV). Pour rappel, une distance de 9.9 Å sépare les deux ammoniums dans  $4_a^{++}$ . Le  $\text{NH}_3$  de la lysine n'étant pas solvée, on s'attend à retrouver certaines propriétés des modèles  $1^\bullet$  et  $2^\bullet$ . En effet, l'énergie verticale de  $\text{ES}^{4a1}$  (0.64 eV) est similaire à la différence des  $\text{ER}_{\text{vert}}$  entre  $1^\bullet$  et  $2^\bullet$  (~ 0.60 eV). Toutefois, cette tendance n'est pas vérifiée pour  $4_b^{++}$  puisqu'une différence de 0.26 eV est estimé tandis que l'écart en énergie entre  $\text{GS}^{4b}$  et  $\text{ES}^{4b3}$  vaut 0.79 eV. L'analyse de la nature des interactions non covalentes avec la méthode NCI montre que dans  $4_a^{2+}$  la partie glycine est peu influencée par la présence de la lysine, contrairement à ce qu'on remarque dans  $4_b^{2+}$  (Figure IV.13). En effet, dans  $4_b^{2+}$  il existe des interactions de van der Waals entre le groupe CO de l'acide carboxylique et les CH (vert) ainsi qu'entre  $\text{HN}^+$  et les liaisons CH (vert) qui stabilise significativement  $4_b^{2+}$ . Lorsque l'électron se localise dans la  $\sigma_{\text{N-H}}^*$  de la lysine ( $\text{ES}^{4b3}$ ), ces interactions de van der Waals deviennent plus faibles déstabilisant cet état excité et augmentant ainsi l'écart en énergie entre  $\text{GS}^{4b}$  et  $\text{ES}^{4b3}$ .

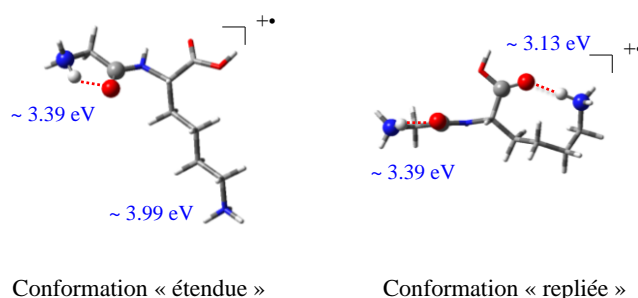


Figure IV.12. Estimation des  $\text{ER}_{\text{vert}}$  LC-BLYP des ammoniums dans  $4_a^{++}$  (à gauche) et  $4_b^{++}$  (à droite) exprimées en eV.

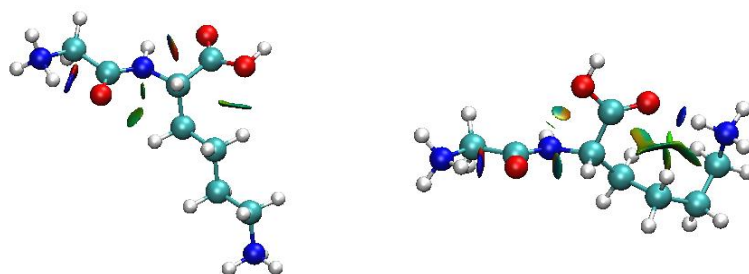


Figure IV.13. Nature des interactions non-covalentes au sein de  $4_a^{2+}$  (à gauche) et  $4_b^{2+}$  (à droite). Les densités SCF ont été utilisées et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.

Par le biais des valeurs des  $ER_{\text{vert}}$  de chaque ammonium du système réel  $5^{+\bullet}$ , nous avons voulu prédire la localisation de la densité de spin à l'état fondamental pour différentes conformations. Dans la conformation  $5_a^{+\bullet}$ , il y a peu de liaisons hydrogène autour de l'ammonium de la glycine. De plus, les valeurs des  $ER_{\text{vert}}$  entre les deux ammoniums sont éloignées puisqu'un écart de 1.33 eV a été calculé (voir Figure IV.14). Il n'y a donc pas d'ambiguïté, l'électron se localise sur le groupement qui possède la plus forte  $ER_{\text{vert}}$ . Dans la conformation  $5_b^{+\bullet}$ , les deux ammoniums sont chacun solvatés deux fois et ont par conséquent des  $ER_{\text{vert}}$  proches. Toutefois, la densité de spin est localisée sur le N-term  $NH_3$  car l'interaction  $NH\cdots OC$  de la glycine est moins forte que  $NH\cdots OCOH$ . Contrairement à  $5_a^{+\bullet}$ , la conformation  $5_c^{+\bullet}$  est plus complexe puisque les deux ammoniums sont solvatés et l' $ER_{\text{vert}}$  des ammoniums séparés est estimé à 2.00 et 1.64 eV pour la glycine et la lysine, respectivement. Il se trouve que l'électron se localise préférentiellement sur un NMA et non pas sur le  $\sigma_{N-H^*}$  disponible, ce qui semble contre-intuitif en considérant les valeurs des  $ER_{\text{vert}}$  seules. Cela confirme que lorsque des environnements locaux de type  $8^\bullet$  et  $9^\bullet$  sont en compétition, le site de capture de l'électron est difficile à prévoir sans effectuer un calcul explicite sur le système complet.

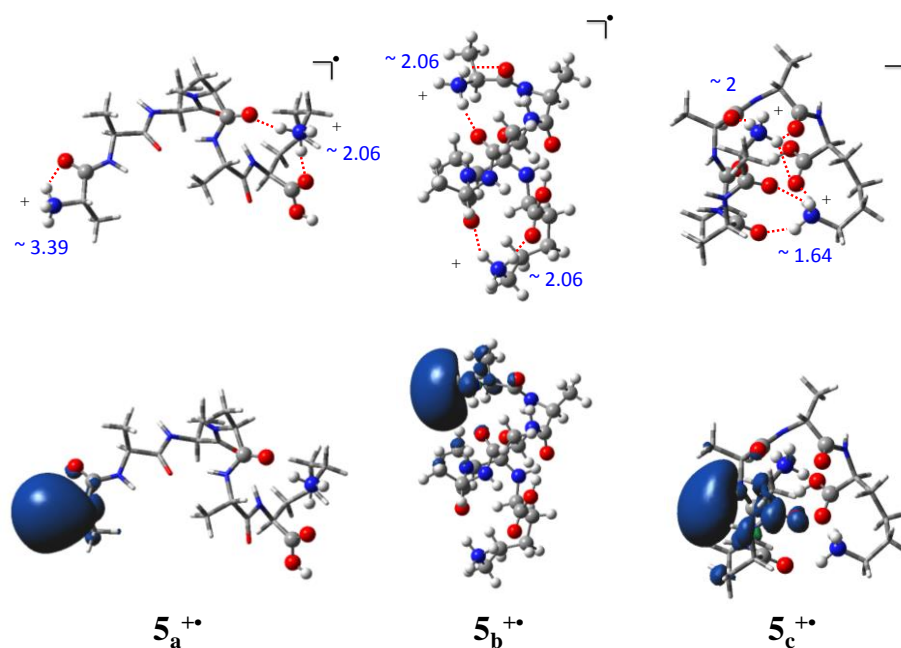


Figure IV.14. Estimation des  $ER_{\text{vert}}$  (exprimées en eV) et densités de spin en LC-BLYP.

## IV.5. Conclusions

L'élucidation du mécanisme de fragmentation de l'intermédiaire radical cation en ECD et ETD passe par une meilleure compréhension de la structure électronique de cet intermédiaire ainsi que par une rationalisation de l'impact des donneurs et accepteurs de liaisons hydrogène sur les différents sites susceptibles de capturer l'électron. Dans le cadre de cette étude, nous avons conçu un banc d'essai de fonctionnelles d'échange-corrélation. La sélection s'est basée sur les méthodes de références CASSCF-MRCI et EOM-CCSD et sur l'utilisation de deux critères (énergies relatives et topologies de la densité de spin dans les différents états électroniques). Nous avons démontré que, en sus d'un échange exact atteignant 100 % à grande distance inter-électronique, il est nécessaire d'utiliser un fort paramètre d'atténuation dans les RSHs pour atteindre une description cohérente avec les méthodes de références. Un faible paramètre d'atténuation ( $\omega$ B97X et  $\omega$ B97X-D) implique une mauvaise description de la structure électronique. Les RSHs LC-BLYP et  $\omega$ B97 sont apparues comme les deux fonctionnelles les plus adaptées pour étudier l'effet des liaisons hydrogène sur la structure électronique des intermédiaires radical cations.

Nous avons mis en évidence l'importance du nombre de liaisons hydrogène solvant l'ammonium captant l'électron. Ces ponts hydrogène influencent le nombre de  $\sigma_{N-H}^*$  disponibles dans l'intermédiaire, ainsi que les énergies de recombinaisons pour les captures dans  $\sigma_{N-H}^*$  et  $\pi_{CO}^*$ . Dans un peptide radical cation contenant plusieurs ammoniums, l'électron se localise préférentiellement dans les  $\sigma_{N-H}^*$  et  $\pi_{CO}^*$  associées à l'ammonium le moins solvato par des carbonyles. Toutefois, si les  $ER_{vert}$  de plusieurs ammoniums sont proches, il devient difficile de prédire sans l'aide du calcul la structure électronique de l'état fondamental de cet intermédiaire.

De ces observations, nous pouvons subodorer que les chemins de réaction tels que la migration ou l'élimination d'un  $H^\bullet$  ou d'un  $H^+$  sont fortement influencés par la présence et le nombre de liaisons hydrogène et cette question mécanistique est traitée dans le chapitre suivant.

## IV.6. Bibliographie

- [1] Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
- [2] McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245–249.
- [3] Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
- [4] Syrstad, E. A.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 208–224.
- [5] Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J. *J. Phys. Chem. B* **2003**, *107*, 13505–13511.
- [6] Tureček, F.; Syrstad, E. A. *J. Am. Chem. Soc.* **2003**, *125*, 3353–3369.
- [7] Siegbahn, P. E. M.; Eriksson, L.; Himo, F.; Pavlov, M. *J. Phys. Chem. B* **1998**, *102*, 10622–10629.
- [8] Cukier, R. I.; Nocera, D. G. *Annu. Rev. Phys. Chem.* **1998**, *49*, 337–369.
- [9] Tureček, F. *Chem. Rev.* **2013**, *113*, 6691–6733.
- [10] Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A. *Int. J. Mass Spectrom.* **2006**, *248*, 124–135.
- [11] Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; Budnik, B. A.; Haselmann, K. F. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1087–1098.
- [12] Adams, C. M.; Zubarev, R. A. *Anal. Chem.* **2005**, *77*, 4571–4580.
- [13] Polfer, N. C.; Haselmann, K. F.; Langridge-Smith, P. R. R.; Barran, P. E. *Mol. Phys.* **2005**, *103*, 1481–1489.
- [14] Van der Rest, G.; Hui, R.; Frison, G.; Chamot-Rooke, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1631–1644.
- [15] Gilson, A. I.; van der Rest, G.; Chamot-Rooke, J.; Kurlancheek, W.; Head-Gordon, M.; Jacquemin, D.; Frison, G. *J. Phys. Chem. Lett.* **2011**, *2*, 1426–1431.
- [16] Jacquemin, D.; Le Bahers, T.; Adamo, C.; Ciofini, I. *Phys. Chem. Chem. Phys.* **2012**, *14*, 555383–5388. Code disponible à l'Université de Nantes, <http://www.sciences.univ-nantes.fr/CEISAM/erc/marches/> (accès Apr. 6, 2014).
- [17] Tureček, F.; Chen, X.; Hao, C. *J. Am. Chem. Soc.* **2008**, *130*, 8818–8833.
- [18] Tureček, F.; Chung, T. W.; Moss, C. L.; Wyer, J. A.; Ehlerding, A.; Holm, A. I. S.; Zettergren, H.; Nielsen, S. B.; Hvelplund, P.; Chamot-Rooke, J.; Bythell, B.; Paizs, B. *J. Am. Chem. Soc.* **2010**, *132*, 10728–10740.
- [19] Chamot-Rooke, J.; Malosse, C.; Frison, G.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2146–2161.
- [20] Frison, G.; van der Rest, G.; Tureček, F.; Besson, T.; Lemaire, J.; Maitre, P.; Chamot-Rooke, J. *J. Am. Chem. Soc.* **2008**, *130*, 14916–14917.
- [21] Panja, S.; Brøndsted Nielson, S.; Hvelplund, P.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1726–1742.
- [22] Jensen, C. S.; Holm, A. I. S.; Zettergren, H.; Overgaard, J. B.; Hvelplund, P.; Brøndsted Nielson, S. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1881–1889.
- [23] Sohn, C. H.; Chung, C. K.; Yin, S.; Ramachandran, P.; Loo, J. A.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2009**, *131*, 5444–5459.
- [24] Bythell, B. *J. Phys. Chem. A* **2013**, *117*, 1189–1196.
- [25] Maclot, S.; Rangama, J.; Brøndsted Nielson, S.; Pouilly J. C. *Int. J. Mass Spectrom.* **2013**, *337*, 1–11.
- [26] Pepin, R.; Laszlo, K. J.; Peng, B.; Marek, A.; Bush, M. F.; Tureček, F. *J. Chem. Phys. A* **2014**, *118*, 308–324.

- [27] Moss, C. L.; Liang, W.; Li, X.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2012**, 23, 446-459.
- [28] Ledvina, A. R.; Chung, T. W.; Hui, R.; Coon, J. J.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2012**, 23, 1351-1363.
- [29] Turecek, F.; Moss, C. L.; Chung, T. W. *Int. J. Mass Spectrom.* **2012**, 330-332, 207-219.

# Chapitre V. Effets des liaisons hydrogène sur des processus de transfert de $H^+$ et/ou de $H^\bullet$



# Partie I. Formation de l'intermédiaire radical cation





## V.1.1. Généralités

Dans le chapitre précédent, nous avons mentionné que dans les mécanismes de Cornell et d'Utah-Washington, le même intermédiaire énol radical cation (Figure V.1) est obtenu dans l'hypothèse où la super base, formée dans ce second mécanisme, ne se fragmente pas directement.

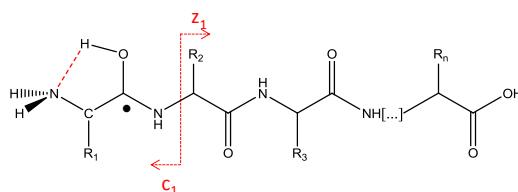


Figure V.1. Intermédiaire radical cation formé lors du mécanisme de Cornell et celui d'Utah-Washington.

Expérimentalement, en plus des ions fragments de type *c* et *z*, il a été observé que la perte d'un radical H est souvent la voie de fragmentation majoritaire obtenue en ECD. De ce fait, plusieurs questions se posent concernant le devenir de l'intermédiaire radical cation après relaxation, notamment si cet intermédiaire est un minima de la surface d'énergie potentielle sans qu'il y ait transfert ou élimination de H<sup>•</sup>. L'existence d'une barrière énergétique pour le transfert de H<sup>•</sup> suggère que cet intermédiaire pourrait exister dans un état stationnaire. En effet, la barrière pour la perte de H<sup>•</sup> pour le système **1**<sup>•</sup> a été calculée à 0.25 [0.22]eV au niveau de théorie CCSD(T)/aug-cc-pVTZ//B3LYP/6-311++G(2d,p) [B3-PMP2/6-311++G(3df,2p)//B3LYP/6-311++G(2d,p)].[1] Cette réaction est très légèrement exothermique, le système formé H<sub>3</sub>CNH<sub>2</sub> + H<sup>•</sup> étant plus stable de -0.05 [-0.1]eV que **1**<sup>•</sup>. Pour **4<sub>a</sub>**<sup>+•</sup>, la barrière associée à ce processus est plus basse (0.16 eV) et la réaction est plus fortement exothermique (-0.21 eV) au niveau B3-PMP2/6-311++G(2d,p)//B3LYP/6-31++G(d,p).[2] Dans cette même étude, une barrière de 0.16 eV a été proposée pour le transfert de H<sup>•</sup> du N-term vers la  $\pi_{CO}^*$  de la glycine pour **4<sub>a</sub>**<sup>+•</sup> considéré dans son état fondamental. Cependant, comme nous l'avons montré dans le Chapitre IV, l'électron n'est pas partiellement localisé sur ce site et ce résultat est peut être lié à la sur-délocalisation de spin entre les deux ammoniums de la méthode de calcul B3-PMP2.

Nous avons noté dans le chapitre précédent que le nombre de liaisons hydrogène environnant l'ammonium modifie de façon non négligeable la répartition spatiale de la densité électronique de l'intermédiaire radical cation. Nous avons donc étudié l'influence du nombre

de liaisons hydrogène autour de  $\text{NH}_3^\bullet$  sur les barrières de transfert et d'élimination d'un  $\text{H}^\bullet$  pour comprendre si ces deux processus sont en compétitions cinétiquement et thermodynamiquement. De même, plusieurs transferts pouvant avoir lieu pour **7** $^\bullet$ , **8** $^\bullet$  et **9** $^\bullet$ , nous avons cherché à estimer si toutes les possibilités de transfert de  $\text{H}^\bullet$  sont envisageables, voire énergétiquement équivalentes. Les transferts de proton loin du site de capture de l'électron ont également été envisagés. En supposant **4** $_{\text{a}}^{+\bullet}$  dans son état fondamental, nous avons cherché différentes voies possibles de transfert de  $\text{H}^+$  et (ou) de  $\text{H}^\bullet$  pour former un intermédiaire énol cation. Les réponses apportées seront des pistes supplémentaires permettant d'élucider le mécanisme de fragmentation en ECD et ETD.

## V.1.2. Etude de la relaxation de peptides radicalux cations

Nous avons étudié l'évolution de la géométrie après réduction. Le Tableau V.1.1 donne, d'une part, les énergies de recombinaisons verticales ( $\text{ER}_{\text{vert}}$ ) des systèmes modèles et, d'autre part, les énergies de relaxation ( $\Delta\text{E}_{\text{rel}}$ ) ; l'énergie de recombinaison totale ( $\text{ER}_{\text{tot}}$ ) étant égale à  $\text{ER}_{\text{vert}} + \Delta\text{E}_{\text{rel}}$ . Les géométries des intermédiaires radicalaires avant la relaxation correspondent à celles obtenues pour les systèmes non-réduits en DFT/6-311++G(2d,p), où nous avons utilisé tant LC-BLYP que  $\omega\text{B97}$ . Cela nous a permis de calculer les  $\text{ER}_{\text{vert}}$ . Ensuite, les géométries des intermédiaires radicalaires ont été optimisées avec la même fonctionnelle permettant ainsi de calculer les  $\Delta\text{E}_{\text{rel}}$ . En considérant l'ammonium non solvaté (**1** $^\bullet$  et **4** $_{\text{a}}^{+\bullet}$ ), on note que leur  $\Delta\text{E}_{\text{rel}}$  sont proches (0.02 et 0.05 eV, respectivement), ce qui est un résultat attendu vu que l'ammonium de la lysine n'est pas vraiment perturbé par le squelette dans **4** $_{\text{a}}^{+\bullet}$ . Comme montré dans le chapitre IV, l'électron est localisé sur l'ammonium de la glycine dans **2** $^\bullet$  et **4** $_{\text{b}}^{+\bullet}$  et sur l'ammonium du NMA dans **7** $^\bullet$ . Ces trois systèmes présentent une liaison hydrogène autour de l'ammonium ayant capté l'électron. **2** $^\bullet$  et **7** $^\bullet$  ont une énergie de relaxation relativement proche (0.15 et 0.12 eV, respectivement) tandis que celle de **4** $_{\text{b}}^{+\bullet}$  est très élevée (1.68 eV). Cette valeur s'explique car un processus de transfert de  $\text{H}^\bullet$  a lieu lors de la relaxation de **4** $_{\text{b}}^{+\bullet}$ , ce qui n'est pas le cas pour les autres systèmes modèles. L'évolution des  $\Delta\text{E}_{\text{rel}}$  de **4** $_{\text{a}}^{+\bullet}$ , **7** $^\bullet$ , **8** $^\bullet$  et **9** $^\bullet$  (0, 1, 2 et 3 liaisons  $\text{H}_2\text{N}^\bullet\text{H}\cdots\text{OC}$ ) indique que la relaxation géométrique augmente avec le nombre de liaisons hydrogène entourant l'ammonium ayant capturé l'électron.

Tableau V.1.1. Énergies de recombinaison verticales ( $ER_{\text{vert}}$ ) et énergies de relaxation ( $\Delta E_{\text{rel}}$ ) associées aux systèmes modèles en eV.

Systèmes	LC-BLYP			$\omega$ B97		
	$ER_{\text{vert}}$	$\Delta E_{\text{rel}}$	$ER_{\text{tot}}$	$ER_{\text{vert}}$	$\Delta E_{\text{rel}}$	$ER_{\text{tot}}$
<b>1<sup>•</sup></b>	4.00	0.02	4.02	3.73	0.02	3.75
<b>2<sup>•</sup></b>	3.33	0.15	3.48	3.12	0.16	3.28
<b>4<sub>a</sub><sup>++</sup></b>	5.29	0.05	5.34	5.01	0.06	5.07
<b>4<sub>b</sub><sup>++</sup></b>	5.19	1.68	6.87	4.94	1.73	6.67
<b>7<sup>•</sup></b>	2.81	0.12	2.93	2.60	0.11	2.71
<b>8<sup>•</sup></b>	2.03	0.20	2.23	1.82	0.20	2.02
<b>9<sup>•</sup></b>	1.62	0.34	1.96	1.41	0.43	1.84

Lors de la capture électronique, l'électron peuple une orbitale de type  $\sigma_{\text{N-H}}^*$  ce qui se traduit par une augmentation des longueurs de liaison N-H non liées ainsi que par une diminution de la longueur N-H liée ( $\text{N-H}\cdots\text{OC}$ ). Cet effet est illustré dans la Figure V.1.2. Pour **7<sup>•</sup>** (**8<sup>•</sup>**), la longueur NH s'allonge de 0.02 (0.03) Å, tandis que  $\text{NH}\cdots\text{OC}$  se contracte de 0.03 (0.01) Å. Ces deux liaisons ne sont pas perturbées lors de la capture électronique dans le système **9<sup>•</sup>** puisque l'électron se localise sur un autre site. De plus, après la relaxation on constate qu'il y a eu un processus de réorganisation assez important pour stabiliser le radical, la densité de spin étant maintenant répartie sur les  $\sigma_{\text{NH}}^*$  de deux NMA.

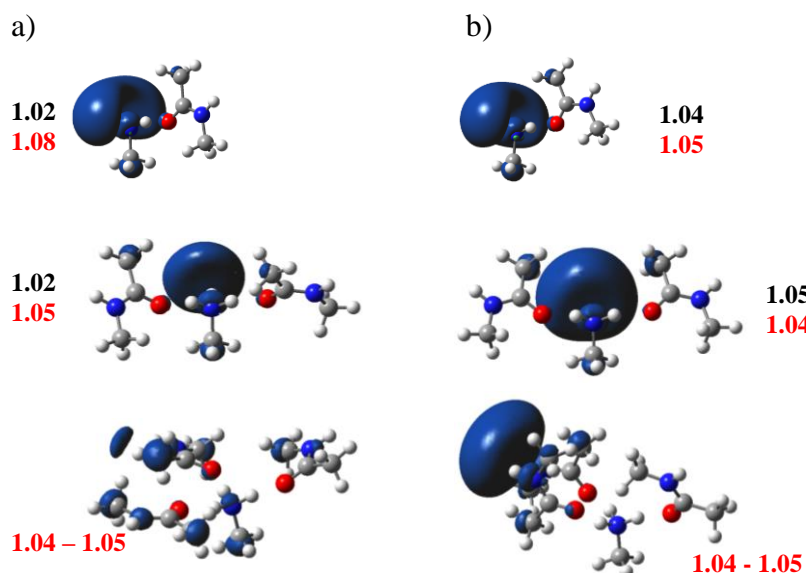


Figure V.1.2. Évolution des longueurs de liaisons N-H non liées (en noir) et liées (en rouge) avec CO a) avant relaxation et b) après relaxation. Les géométries sont celles obtenues au niveau LC-BLYP et les distances sont en Å.

Si le transfert de  $\text{H}^{\bullet}$  semble s'effectuer sans barrière lors de la relaxation de **4<sub>b</sub><sup>++</sup>** (cf ci-dessous), il ne s'initie pas spontanément pour **2<sup>•</sup>** pour lequel il est nécessaire de franchir une

barrière d'activation de 0.23 eV (Figure V.1.3). L'analyse de la densité de spin de l'état de transition  $\text{TS}_\text{M}2^\bullet$  nous indique que tout au long du chemin, la densité de spin n'est jamais localisée sur le H qui migre. Ce phénomène est également présent pour les états de transition associés aux transferts de  $\text{H}^\bullet$  vers la  $\pi_{\text{CO}}^*$  des systèmes  $7^\bullet$  et  $8^\bullet$  (partie a) de la Figure V.1.4). Ainsi, il y a un transfert couplé proton-électron simultané de l'ammonium vers  $\pi_{\text{CO}}^*$  de la glycine, comme cela a déjà été suggéré dans la littérature.[3-6] Concernant le processus d'élimination, il s'agit bien d'un radical H qui s'élimine comme l'atteste la densité de spin de cet atome dans les états de transition représentés dans la partie b) de la Figure V.1.4 pour les composés  $1^\bullet$ ,  $2^\bullet$ ,  $7^\bullet$  et  $8^\bullet$ .

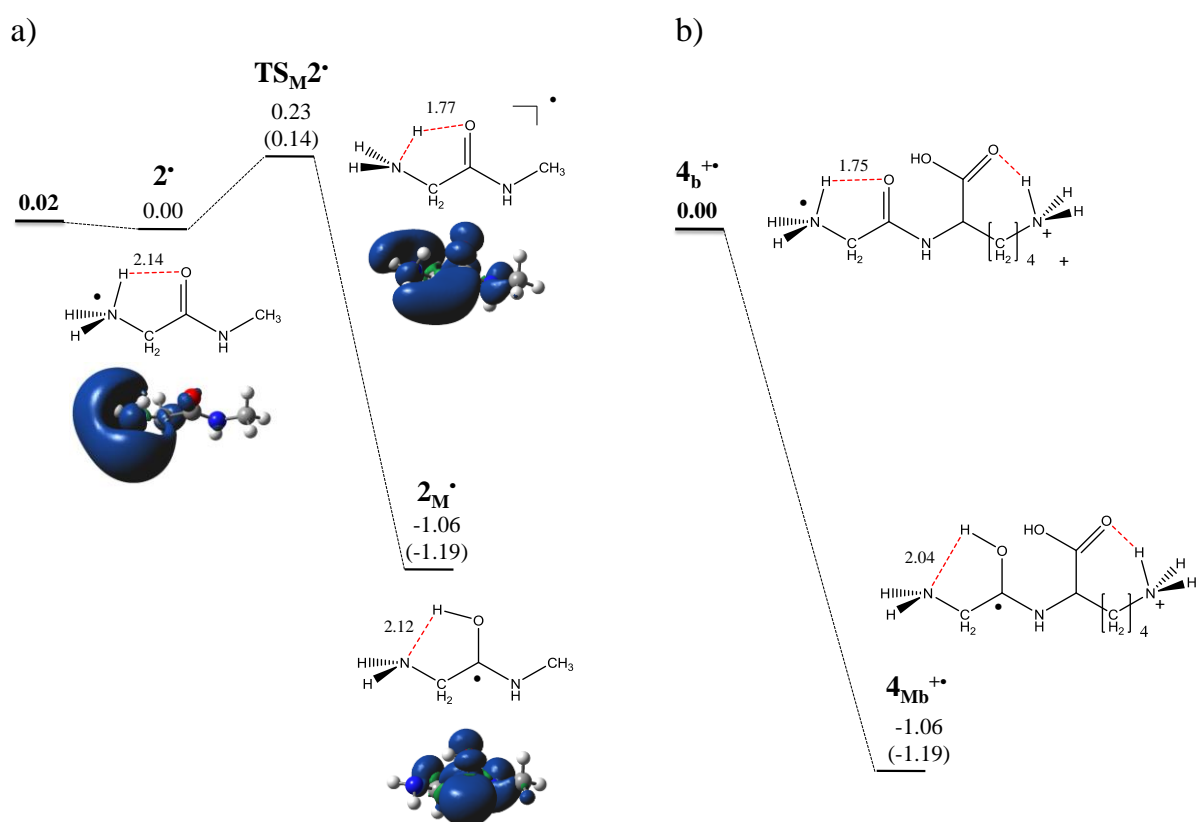


Figure V.1.3. Effet de la charge sur l'évolution de la barrière correspondant au transfert couplé proton-électron pour les systèmes a)  $2^\bullet$  et b)  $4_\text{b}^{+\bullet}$ . Les énergies relatives LC-BLYP ( $\omega$ B97) sont en eV.

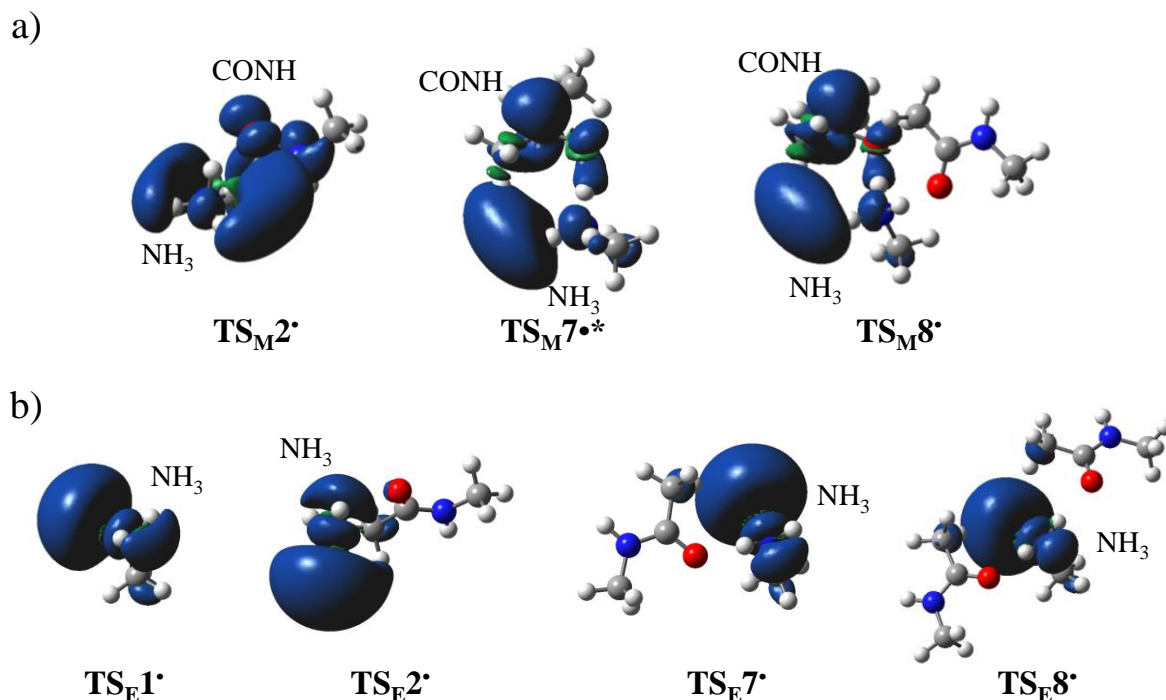


Figure V.1.4. Densité de spin des états de transitions associés a) au transfert couplé proton-électron et b) à l'élimination d'un H• (géométries LC-BLYP). \* géométrie ωB97.

Afin de comprendre pourquoi le transfert simultané d'un proton et d'un électron s'effectue lors de la relaxation de  $4_b^{+•}$ , nous avons comparé les densités de spin des structures  $2^•$  et  $4_b^{+•}$  avant relaxation (Figure V.1.5). Avant la relaxation, la délocalisation de spin a déjà été initiée dans les systèmes  $2^•$  et  $4_b^{+•}$ , mais elle est plus importante dans  $4_b^{+•}$  (0.27) que dans  $2^•$  (0.12), ce qui résulte probablement d'un effet de charge. Cette délocalisation déjà « avancée » permet d'abaisser voire de faire disparaître la barrière d'activation.

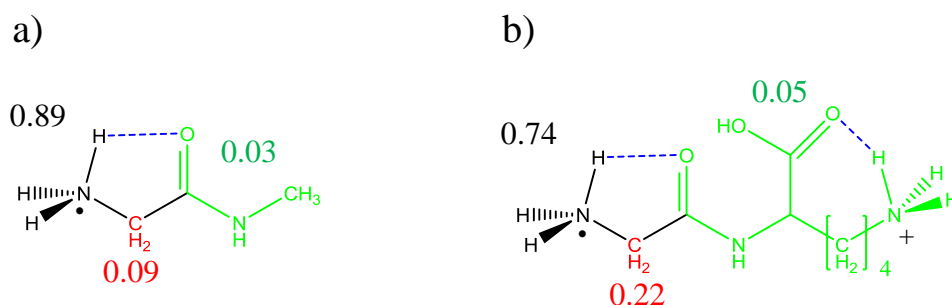


Figure V.1.5. Répartition des densités de spin atomiques des groupements NH<sub>3</sub> (en noir), CH<sub>2</sub> (en rouge) et OCNHCH<sub>3</sub> (en vert) pour a)  $2^•$  et b)  $4_b^{+•}$  avant relaxation.

Nous avons reporté dans le Tableau V.1.2 les énergies, enthalpies et enthalpies libres relatives, associées aux processus d'élimination de H• et de transfert couplé proton-électron.

Ces Tables donnent également les barrières d'activation et des systèmes formés. La perte et la migration d'un radical H sont des processus exothermiques, mais l'élimination de H<sup>•</sup> dégage plus d'énergie.

Tableau V.1.2. Énergies relatives LC-BLYP (ωB97) des états de transition et de l'intermédiaire radical cation dissocié en eV.\*

Modèles	Systèmes	Élimination H <sup>•</sup>			Systèmes	Migration H <sup>•</sup>		
		ΔE°	ΔH° <sub>298</sub>	ΔG° <sub>298</sub>		ΔE°	ΔH° <sub>298</sub>	ΔG° <sub>298</sub>
1 <sup>•</sup>	TS <sub>E</sub> 1 <sup>•</sup>	0.31 (0.29)	0.12 (0.11)	0.08 (0.06)	TS <sub>M</sub> 1 <sup>•</sup>	/	/	/
	1 <sub>E</sub> <sup>•</sup>	0.06 (-0.21)	-0.16 (-0.42)	-0.51 (-0.77)	1 <sub>M</sub> <sup>•</sup>	/	/	/
2 <sup>•</sup>	TS <sub>E</sub> 2 <sup>•</sup>	0.17 (0.16)	0.00 (0.00)	-0.02 (-0.03)	TS <sub>M</sub> 2 <sup>•</sup>	0.23 (0.14)	0.20 (0.08)	0.20 (0.12)
	2 <sub>E</sub> <sup>•</sup>	-0.12 (-0.41)	-0.32 (-0.60)	-0.64 (-0.98)	2 <sub>M</sub> <sup>•</sup>	-1.06 (-1.19)	-1.02 (-1.13)	-0.98 (-1.11)
4 <sub>a</sub> <sup>++</sup>	TS <sub>E1</sub> 4 <sub>a</sub> <sup>++</sup>	0.26 (0.24)	0.09 (0.08)	0.07 (0.06)	TS <sub>M1</sub> 4 <sub>a</sub> <sup>++</sup>	/	/	/
	4 <sub>E1a</sub> <sup>++</sup>	-0.15 (-0.42)	-0.36 (-0.61)	-0.69 (-0.94)	4 <sub>M1a</sub> <sup>++</sup>	/	/	/
	TS <sub>E2</sub> 4 <sub>a</sub> <sup>++</sup>	0.16 (0.15)	0.00 (0.00)	0.00 (-0.01)	TS <sub>M2</sub> 4 <sub>a</sub> <sup>++</sup>	/	/	/
	4 <sub>E2a</sub> <sup>++</sup>	-0.19 (-0.46)	-0.40 (-0.65)	-0.72 (-0.98)	4 <sub>M2a</sub> <sup>++</sup>	/	/	/
	TS <sub>E3</sub> 4 <sub>a</sub> <sup>++</sup>	0.19 (0.17)	0.03 (0.02)	0.03 (0.02)	TS <sub>M3</sub> 4 <sub>a</sub> <sup>++</sup>	/	/	/
	4 <sub>E3a</sub> <sup>++</sup>	-0.19 (-0.46)	-0.39 (-0.65)	-0.72 (-0.98)	4 <sub>M3a</sub> <sup>++</sup>	/	/	/
7 <sup>•</sup>	TS <sub>E</sub> 7 <sup>•</sup>	0.29 (0.24)	0.08 (0.03)	0.09 (0.06)	TS <sub>M</sub> 7 <sup>•</sup>	/ (0.37)	/ (0.21)	/ (0.32)
	7 <sub>E</sub> <sup>•</sup>	0.21 (-0.12)	-0.03 (-0.37)	-0.38 (-0.65)	7 <sub>M</sub> <sup>•</sup>	/ (-0.90)	/ (-0.90)	/ (-0.89)
8 <sup>•</sup>	TS <sub>E</sub> 8 <sup>•</sup>	0.32 (0.24)	0.09 (0.01)	0.06 (-0.03)	TS <sub>M</sub> 8 <sup>•</sup>	0.21 (0.17)	0.08 (0.03)	0.18 (0.07)
	8 <sub>E</sub> <sup>•</sup>	0.39 (0.02)	0.12 (-0.25)	-0.27 (-0.71)	8 <sub>M</sub> <sup>•</sup>	-0.80 (-0.96)	-0.83 (-0.99)	-0.82 (-0.98)
		Élimination de H <sup>+</sup>					Migration de H <sup>+</sup>	
4 <sub>a</sub> <sup>++</sup>	TS <sub>E4</sub> 4 <sub>a</sub> <sup>++</sup>	/	/	/	TS <sub>M4</sub> 4 <sub>a</sub> <sup>++</sup>	0.09 (0.18)	0.02 (0.05)	0.01 (0.09)
	4 <sub>E4a</sub> <sup>++</sup>	/	/	/	4 <sub>M4a</sub> <sup>++</sup>	0.00 (0.04)	-0.03 (0.01)	-0.02 (0.03)

\*Les valeurs manquantes indiquent que les processus ne sont pas possibles.

La barrière d'activation pour l'élimination d'un  $H^\bullet$  est indépendante du nombre de liaisons hydrogène autour de l'ammonium ayant capté l'électron. En effet,  $TS_E1$ ,  $TS_E7$  et  $TS_E8$  implique 0, 1 et 2 liaisons hydrogène autour de  $NH_3^\bullet$  et leurs énergies relatives sont 0.31, 0.29 et 0.32 eV, respectivement. L'élimination de  $H^\bullet$  n'est pas équivalente pour les trois atomes d'hydrogène de  $NH_3$  de la lysine dans  $4_a^{+\bullet}$ , celui qui s'élimine en direction opposée du squelette peptidique présentant une barrière plus importante (Tableau V.1.2 et partie a) de la Figure V.1.6). Il appert que le squelette peptidique facilite le processus, se traduisant par un abaissement de la barrière de  $\sim 0.10$  eV. En considérant les contributions vibrationnelles et entropiques, les barrières d'activation des systèmes modèles deviennent toutes inférieures à 0.1eV.

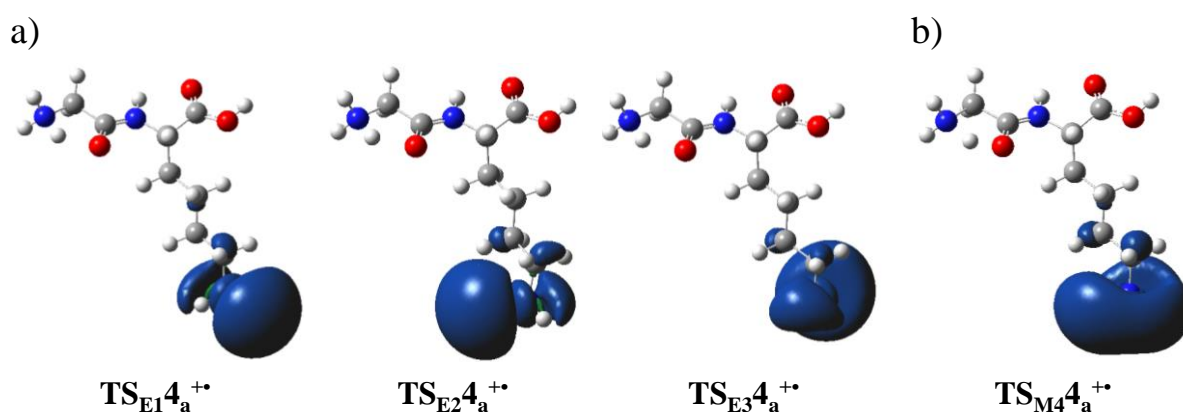


Figure V.1.6. Densité de spin des états de transitions du modèle  $4_a^{+\bullet}$  associés a) à l'élimination d'un  $H^\bullet$  et b) à la migration d'un  $H^+$  (géométries LC-BLYP).

Dans le cas du processus du transfert couplé proton-électron, les barrières en énergie totale et en enthalpie libre sont proches. Nous n'avons pas identifié  $TS_M7$  avec le niveau de calcul LC-BLYP, donc exceptionnellement, nous commenterons les données  $\omega B97$ . Les liaisons hydrogène environnantes le groupement  $NH_3^\bullet$  permettent de favoriser la migration couplée proton-électron vers l'orbitale  $\pi_{CO}^*$ . En effet, la barrière d'activation en enthalpie libre s'abaisse lorsque le nombre ou la force des liaisons H augmentent (pour  $TS_M7$ ,  $TS_M2$  et  $TS_M8$  elles valent 0.32, 0.12 et 0.07 eV).

La migration d'un proton vers la  $\pi_{CO}^*$  ( $TS_{M4}4_a^{+\bullet}$  dans la partie b) de la Figure V.1.6) présente une barrière d'activation quasi inexistante (0.01 eV en LC-BLYP).

Comme remarqué précédemment, l'existence de la barrière d'activation pour le transfert de  $H^+ + e^-$  dépend fortement des groupements à proximité de l'ammonium ayant capté l'électron, et la présence ainsi que son amplitude doivent être déterminées. Toutefois, si elle existe, cette



barrière d'activation reste, selon nos modèles, nettement inférieure à 0.5 eV. Cela suggère que l'intermédiaire radical cation sans transfert ou élimination de H<sup>•</sup> peut être un point stationnaire de la surface d'énergie potentielle. Lors de la recombinaison électronique, le système va emmagasiner de l'énergie qui est nettement supérieure à cette barrière d'activation. La plus faible valeur calculée pour cette recombinaison est de 1.62 eV pour **9'** (sans considérer la relaxation). Ainsi, le passage de la barrière d'activation se fait très facilement.

Un raisonnement similaire peut être effectué pour l'élimination de H<sup>•</sup> qui passe par des barrières d'activation encore plus faible (< 0.2 eV en enthalpie libre) que celles de la migration. Cela permet de justifier également que la perte d'un radical H est souvent la voie de fragmentation majoritaire obtenue lors des expériences en ECD.

Pour le système **4<sub>a</sub><sup>++</sup>**, nous n'avons pas identifié d'état de transition permettant d'observer le transfert de H<sup>+</sup> + e<sup>-</sup> vers la  $\pi_{CO}^*$  (voir Figure V.17). Cela n'est pas surprenant car l'électron est localisé sur la  $\sigma_{N-H}^*$  de la lysine. Toutefois, la localisation d'un radical H sur la  $\pi_{CO}^*$  dans le système **4<sub>a</sub><sup>++</sup>** conduit à la formation d'un intermédiaire plus stable de -1.08 eV. Cet intermédiaire peut provenir de plusieurs chemins différents. Une première hypothèse est que la capture de l'électron conduise directement ou non à l'intermédiaire radical cation dans son état fondamental. La formation de l'intermédiaire énol radical cation résulterait d'abord d'un transfert de H<sup>+</sup> qui a une faible barrière d'activation (0.09 eV) suivi d'un transfert de l'électron de NH<sub>3</sub> de la lysine vers la  $\pi_{CO}^*$ . Une seconde hypothèse est que la capture de l'électron par **4<sub>a</sub><sup>++</sup>** dans la glycine conduise à un intermédiaire situé dans l'un de ces états excités, i.e ES<sup>4a1</sup> et ES<sup>4a5</sup> par exemple. Lorsque l'intermédiaire formé après capture de l'électron se trouve dans l'état ES<sup>4a1</sup>, ce système pourrait se désexciter par le transfert de H<sup>+</sup> + e<sup>-</sup> vers la  $\pi_{CO}^*$  (mécanisme type Cornell). Concernant le mécanisme d'Utah-Washington, si l'électron se retrouve dans l'état excité ES<sup>4a5</sup>, un transfert de proton vers la  $\pi_{CO}^*$  conduirait à la formation de l'intermédiaire radical énol.

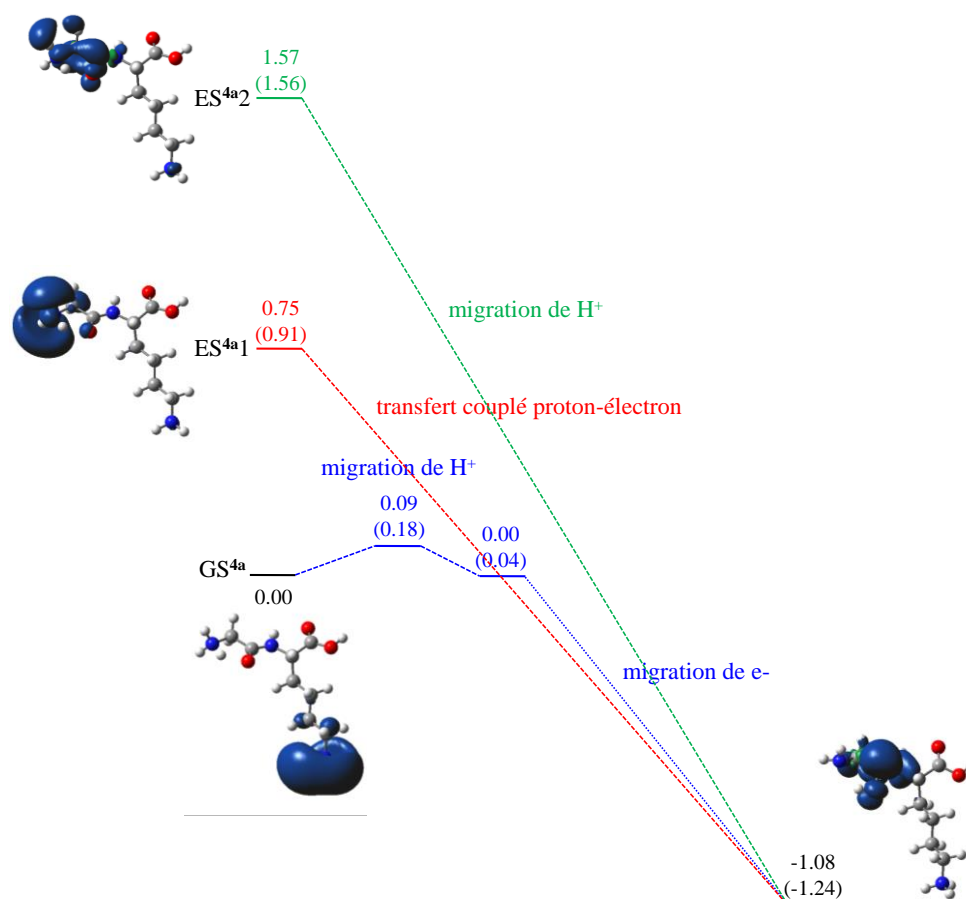


Figure V.1.7. Voies possibles pour la formation de l'intermédiaire émol radical cation. Énergies relatives (eV) obtenues en LC-BLYP ( $\omega$ B97).

### V.1.3. Conclusions

L'étude de la relaxation nous a permis de répondre à plusieurs questions posées dans la partie V.1.1. Au cours de la relaxation du système, la relaxation géométrique augmente avec le nombre de liaisons hydrogène entourant l'ammonium afin de pouvoir stabiliser le radical. En revanche, les sites ne capturant pas l'électron sont peu influencés par la présence de l'électron célibataire comme l'attestent les longueurs de liaisons N-H dans le système **9**<sup>•</sup> non perturbés par la présence du radical dans la  $\sigma_{\text{N-H}}^*$  du NMA.

Le transfert couplé proton-électron et l'élimination de H<sup>•</sup> sont deux processus possédant de faibles barrières ( $\Delta G^\circ_{298} < 0.5$  eV et  $\sim 0.1$  eV, respectivement). Toutefois, l'élimination d'un radical H est favorisée d'un point de vue cinétique et thermodynamique par rapport au transfert couplé justifiant ainsi que ce premier processus soit majoritaire expérimentalement. Le squelette peptidique influençant l'élimination de H<sup>•</sup>, il peut cependant rendre ce second processus compétitif.

L'intermédiaire radical cation n'est pas nécessairement un point stationnaire de la surface d'énergie potentielle. En effet, un effet de charge peut initier la délocalisation de la densité de spin de  $\sigma_{\text{N-H}}^*$  vers la  $\pi_{\text{CO}}^*$  lors de la relaxation, abaissant ou éliminant la barrière d'activation associée au processus de transfert de  $\text{H}^\bullet$ . Cela conduit à la formation directe de l'intermédiaire énoï radical cation, un intermédiaire postulé dans les mécanismes de Cornell et d'Utah-Washington.

Dans le cas de  $4_{\text{a}}^{+*}$ , plusieurs voies sont possibles pour former l'énol radical cation. L'une d'elle implique un transfert de proton suivi d'un transfert d'électron lorsque l'intermédiaire radical cation est dans son état fondamental. Les premiers états excités dans lesquels l'électron est localisé, soit dans la  $\sigma_{\text{N-H}}^*$  soit dans la  $\pi_{\text{CO}}^*$ , permettent également de former cet énoï. Les processus associés sont un transfert couplé proton-électron et une migration de proton, respectivement. Ainsi, le mécanisme de Cornell peut également avoir lieu avec l'un des états excités de l'intermédiaire radical cation.

## V.1.4. Bibliographie

- [1] Yao, C.; Tureček, F. *Phys. Chem. Chem. Phys.* **2005**, 7, 912-920.
- [2] Tureček, F.; Chen, X.; Hao, C. *J. Am. Chem. Soc.* **2008**, 130, 8818-8833.
- [3] Tureček, F.; Syrstad, E. A. *J. Am. Chem. Soc.* **2003**, 125, 3353-3369.
- [4] Siegbahn, P. E. M.; Eriksson, L.; Himo, F.; Pavlov, M. *J. Phys. Chem. B* **1998**, 102, 10622-10629.
- [5] Cukier, R. I.; Nocera, D. G. *Annu. Rev. Phys. Chem.* **1998**, 49, 337-369.
- [6] Tureček, F.; Julian, R. R. *Chem. Rev.* **2013**, 113, 6691-6733.



Partie II. Effet de la micro-solvatation sur  
les premières étapes de la réaction de  
Strecker. Exemple de la glycine



Les acides aminés constituent l'une des plus importantes classes de composés naturels. La participation d'une vingtaine d'acides  $\alpha$ -aminés à la construction des protéines du vivant n'est pas le moindre de leur intérêt. Leurs modes de synthèses, par voies chimiques ou biochimiques, ont en conséquence fait l'objet de nombreuses études.[1,2] Parmi les plus anciennes de ces méthodes, la réaction de Strecker[3] est toujours utilisée en chimie de synthèse [4-7] et son hypothétique intervention dans la genèse prébiotique des acides aminés est abondamment évoquée [8-15].

La première étape de la réaction de Strecker est une addition d'ammoniac sur un aldéhyde (Figure V.2.1). L'aminoolcool résultant de cette addition peut se déshydrater pour former une imine. A ce stade, l'addition de cyanure d'hydrogène sur l'imine permet de former un aminonitrile qui par hydrolyse conduit à la formation de l'acide aminé correspondant.

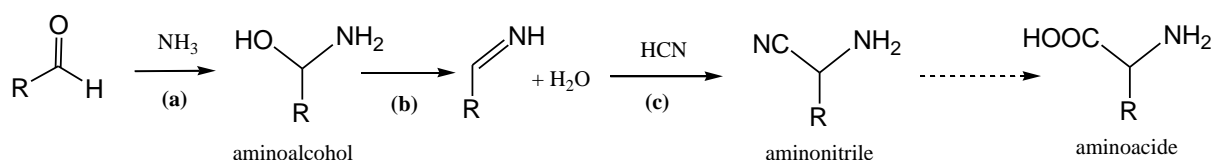


Figure V.2.1. La synthèse de Strecker.

C'est avec comme objectif d'examiner l'influence de l'hydratation sur la faisabilité de la réaction de Strecker, que nous avons entrepris d'examiner les étapes (a) et (b) (Figure V.2.1) dans le cas simple  $R=H$ . Avant d'aborder cette voie de synthèse de la glycine, nous présenterons le contexte qui suggère la possibilité d'une telle réaction dans les conditions prébiotiques.

## V.2.1. La chimie prébiotique

Les molécules du vivant sont principalement (i) les acides nucléiques responsables du codage et du transport de l'information génétique, (ii) les protéines qui contribuent aux processus métaboliques par leurs rôles catalyseurs, régulateurs et coordinateurs, et (iii) les lipides qui forment les membranes cellulaires et participent aux échanges qui s'y déroulent. L'hypothèse de base en chimie prébiotique consiste à considérer que ces molécules complexes ont été formées à partir de constituants plus petits : bases puriques et pyrimidiques, acides aminés et acides gras. Ces "briques élémentaires" seraient elles-mêmes issues de matériaux initiaux constitués de quelques atomes seulement :  $H_2$ ,  $CH_4$ ,  $NH_3$ ,  $HCN$ ,  $CO$ ,  $CO_2$ ,  $H_2O$ ,  $CH_2O$ ... et



formées soit de manière endogène (i.e. sur la Terre primitive) soit de manière exogène (i.e. plus loin dans le système solaire et dans le milieu interstellaire).

Les hypothèses de formation endogène de molécules prébiotiques reposent sur des modèles d'atmosphère et d'océans planétaires primitifs. En 1953, Stanley Miller utilisa un mélange gazeux composé de  $\text{CH}_4$ ,  $\text{NH}_3$ ,  $\text{H}_2\text{O}$  et  $\text{H}_2$ . [9] Des décharges électriques simulant des éclairs appliquées à travers ce mélange gazeux ont conduit à divers produits dont des acides aminés (mais aussi de l'acide aspartique et de l'acide  $\alpha$ -amino-n-butyrique). Plus récemment, des expériences analogues ont été réalisées en utilisant une atmosphère moins réductrice, et semble-t-il plus réaliste, constituée de  $\text{H}_2\text{O}$ ,  $\text{CO}_2$  et  $\text{N}_2$ . À l'issue de ces expériences, des acides aminés ont également été détectés. [16-18] Il est intéressant de noter que des molécules telles que  $\text{CH}_2\text{O}$ ,  $\text{HCN}$  (i.e. les réactifs de la réaction de Strecker) sont également formées lors de ces expériences (Noter aussi que ces petites molécules sont les précurseurs des sucres et des purines). La formation de molécules prébiotiques autour des sources hydrothermales des dorsales océaniques a également été proposée. Des expériences de simulations menées en laboratoire, ont permis en effet de synthétiser des acides aminés à partir de molécules inorganiques dans un milieu aqueux. Les plus hauts rendements ont été obtenus pour des températures de l'ordre des 100-200 °C. [19]

La quantité de matière organique ayant atteint la surface de la Terre durant les 100 millions d'années situées entre -3.9 et -3.8 milliards d'années, est estimée aux alentours de 100 fois la quantité présente actuellement. [20] En conséquence, la formation exogène de molécules prébiotique puis leur dépôt sur la Terre primitive est un scénario raisonnable. Les analyses de météorites (chondrite, [21-23] Murchison, [24] Murray, [25]) et de comètes (prélèvements par sondes spatiales [26]) révèlent en effet la présence d'acides aminés. De plus, la présence de sous-produits habituels de la synthèse de Strecker (acides imino dicarboxyliques en particulier) dans les météorites chondrites a également été mise en évidence. [12] Enfin, les signatures spectrales dans les domaines UV-visible, IR et radiofréquence des précurseurs élémentaires des acides aminés (cyanure d'hydrogène, ammoniac, formaldéhyde et l'acétaldéhyde) ou de molécules proches (aldéhydes, acides, amides, urées, cyanides) ont été identifiées au sein des nuages interstellaires (une liste non exhaustive des molécules détectées est présentée dans le Tableau V.2.1).

Tableau V.2.1. Molécules, ions et radicaux détectés par les méthodes spectroscopiques dans le milieu interstellaire.<sup>1</sup>

Composés hydrogénés						
H <sub>2</sub>			H <sub>3</sub> <sup>+</sup>			
Composés contenant du carbone, de l'hydrogène et de l'oxygène						
OH	CO	CO <sup>+</sup>	H <sub>2</sub> O	HCO	HCO <sup>+</sup>	HOC <sup>+</sup>
C <sub>2</sub> O	CO <sub>2</sub>	H <sub>3</sub> O <sup>+</sup>	HOCO <sup>+</sup>	H <sub>2</sub> CO	C <sub>3</sub> O	HCOOH
H <sub>2</sub> CCO	H <sub>2</sub> COH <sup>+</sup>	H <sub>3</sub> COH	H <sub>2</sub> CCHO	HC <sub>2</sub> CHO	C <sub>5</sub> O	H <sub>3</sub> CCHO
c-C <sub>2</sub> H <sub>4</sub> O	H <sub>3</sub> COCH	H <sub>2</sub> COHCH	H <sub>3</sub> CCOOH	H <sub>2</sub> CCHO	H <sub>3</sub> COC	H <sub>2</sub> CCHCH
	O	O		H	H <sub>3</sub>	O
<sub>3</sub> HCCH <sub>2</sub> CH	H <sub>3</sub> CCH <sub>2</sub> O	H <sub>3</sub> CCOCH	HOCH <sub>2</sub> CH <sub>2</sub> O	C <sub>2</sub> H <sub>5</sub> OC		
O	H	<sub>3</sub>	H	H <sub>3</sub>		
Composés contenant de l'azote, de l'hydrogène et du carbone						
NH	CN	NH <sub>2</sub>	HCN	HNC	N <sub>2</sub> H <sup>+</sup>	NH <sub>3</sub>
HCNH <sup>+</sup>	H <sub>2</sub> CN	HCCN	C <sub>3</sub> N	H <sub>2</sub> CCN	H <sub>2</sub> CNH	HC <sub>3</sub> N
HC <sub>2</sub> NC	H <sub>2</sub> NCN	C <sub>3</sub> NH	H <sub>3</sub> CCN	H <sub>3</sub> CNC	HC <sub>3</sub> NH <sup>+</sup>	C <sub>5</sub> N
H <sub>3</sub> CNH <sub>2</sub>	C <sub>2</sub> H <sub>3</sub> CN	HC <sub>5</sub> N	H <sub>3</sub> CC <sub>3</sub> N	C <sub>2</sub> H <sub>5</sub> CN	HC <sub>7</sub> N	HC <sub>9</sub> N
HC <sub>11</sub> N						
Composés contenant de l'azote, du carbone, de l'hydrogène et de l'oxygène						
NO	HNO	N <sub>2</sub> O	HNCO	H <sub>2</sub> NCHO		
Composés soufrés, silicés et autres espèces autres que deutérées, chaînes et cycles carbonés						
SH	CS	SO	SO <sup>+</sup>	NS	SiH	SiC
SiN	SiO	SiS	HCl	NaCl	AlCl	KCl
HF	AlF	CP	PN	H <sub>2</sub> S	C <sub>2</sub> S	SO <sub>2</sub>
OCS	HCS <sup>+</sup>	c-SiC <sub>2</sub>	SiCN	NaCN	MgCN	MgNC
H <sub>2</sub> CS	HNCS	C <sub>3</sub> S	c-SiC <sub>3</sub>	SiH <sub>4</sub>	SiC <sub>4</sub>	H <sub>3</sub> CSH
C <sub>5</sub> S	FeO	AlNC				

<sup>1</sup> Données issues du site <http://www.obs.u-bordeaux1.fr/pcmi/molecules/table-mol.html> (Programme National de Physique Chimie du Milieu Interstellaire) et de la référence [27].

L'hypothèse "exogène" suppose une synthèse des molécules élémentaires au sein du milieu interstellaire. Ce dernier est constitué d'un mélange non homogène de gaz et de poussières. On distingue, (i) les nuages interstellaires "diffus", constitué essentiellement d'hydrogène atomique et moléculaire et présentant une faible densité de particules ( $\sim 100$  particules.cm<sup>-3</sup>) et des températures de l'ordre de 100-200 K, et (ii) les nuages plus denses, dits « nuages moléculaires », constitués de grains de poussières dont la densité augmente lorsqu'on se rapproche du cœur du nuage. Globalement, les nuages moléculaires présentent une densité située entre 10<sup>2</sup>-10<sup>6</sup> particules.cm<sup>-3</sup>. De plus, leur température est plus faible (5-20 K) que dans les nuages diffus.[28] Les grains de poussières ( $\sim 0.1$  μm) contenus dans ces nuages sont principalement constitués de silicates et de composés carbonés. Par accrétion ces grains se recouvrent d'une couche de glace "sale" (contenant H<sub>2</sub>O, NH<sub>3</sub>, CO, CO<sub>2</sub>, ... notamment) qui peut être le siège de réactions chimiques soit dans la phase condensée soit dans la phase gazeuse, lors de la sublimation des enveloppes de glaces, soit à l'interface.

## V.2.2. La synthèse de Strecker: une voie de formation prébiotique possible pour les acides aminés?

Afin de savoir si la réaction de Strecker est réalisable dans les conditions allant de la phase gazeuse à la phase adsorbée sur une surface de glace ou immergée dans l'eau, une étude mécanistique et énergétique des étapes **(a)** et **(b)** de la Figure V.2.1 est nécessaire. Quelques études théoriques ont été publiées antérieurement afin d'éclairer cette question. Leurs conclusions peuvent être résumées de la manière suivante :

- 1) l'énergie critique (à 0 K) de la barrière associée à l'étape **(a)** a été située entre 115 et 155 kJ mol<sup>-1</sup> pour le système H<sub>2</sub>CO + NH<sub>3</sub> selon le niveau de théorie utilisé (B3LYP, MP2, MP4, CCSD, CCSD(T)).[29-34]
- 2) L'énergie critique de l'étape **(b)** est plus haute que celle de l'étape **(a)** de ~ 50-90 kJ mol<sup>-1</sup> en fonction du niveau de théorie utilisé.[30,34]
- 3) L'effet de l'environnement a été mimé soit par un modèle polarisable de surface d'isodensité, noté IPCM, sans prendre en compte la modification de la géométrie due aux effets du solvant [29] soit par quelques molécules de H<sub>2</sub>O explicites et actives au cours de la catalyse.[29,32,34] Les valeurs obtenues lors de ces diverses études sont assez disparates et il est difficile d'en tirer des conclusions cohérentes. On constate par exemple que la méthode de calcul utilisée a plus d'effet sur les énergies critiques des étapes **(a)** et **(b)** que le nombre de molécules d'eau introduites dans le système. Une exploration plus systématique, menée à un niveau de calcul élevé semble donc nécessaire, c'est ce que nous proposons ci-après.

Il apparaît par conséquent que les aspects mécanistiques et énergétiques des étapes **(a)** et **(b)** de la réaction de Strecker ne sont pas encore solidement étayés. Les études théoriques antérieures ne comportent pas de recherches exhaustives des divers minima et tous les types de catalyses actives ou passives par quelques molécules d'eau n'ont pas été envisagés. De plus, une étude à un haut niveau de théorie devrait être envisagée pour calculer de manière précise les barrières d'énergie afin de permettre des comparaisons sérieuses. Une optimisation de géométrie dans le solvant est également cruciale. Pour aborder ces questions, nous avons examiné les deux premières étapes, **(a)** et **(b)**, de la synthèse de Strecker de la glycine par des

calculs de chimie quantique menés jusqu'au niveau de théorie G3B3. Nous avons d'abord étudié ces deux étapes en phase gazeuse avant d'ajouter progressivement des molécules de H<sub>2</sub>O explicites, puis nous avons utilisé le modèle de solvant C-PCM pour mimer l'effet global du solvant ou de la glace. L'influence du milieu (neutre ou protoné) a également été considérée.

## V.2.3. Stratégies

Pour construire les profils enthalpie/trajet réactionnel des divers systèmes étudiés, nous avons recherché de manière exhaustive les structures des complexes d'approche et des intermédiaires aminométhanol diversement hydratés. Nous avons pour cela utilisé notre stratégie de recherches de minimas développée au chapitre III partie 1 en associant approche hiérarchique et arbre de Darwin.

### Recherches des structures

La première tâche a été de rechercher les structures de plus basses énergies de l'aminométhanol *n*-hydraté neutre et protoné : HOCH<sub>2</sub>NH<sub>2</sub>(H<sub>2</sub>O)<sub>*n*</sub> et HOCH<sub>2</sub>NH<sub>3</sub><sup>+</sup>(H<sub>2</sub>O)<sub>*n*</sub> respectivement avec *n* = 0 à 4. Pour *n* = 0 à 2 l'intuition chimique a été suffisante compte tenu de la faible taille des systèmes. En revanche, pour *n* = 3 et 4 les diverses structures ont été recherchées par des simulations de dynamique moléculaire avec le champ de force polarisable AMOEBA. Pour cela, nous avons dû construire les paramètres de ces deux molécules qui ne sont pas disponibles tels quels dans le fichier de paramètres AMOEBA. Les paramètres de l'éthylamine (NH<sub>2</sub>CH<sub>2</sub>-) et de l'éthanol (-CH<sub>2</sub>OH) nous ont permis de construire les paramètres de l'aminométhanol ; et ceux de l'aminométhanol protoné ont été construits à partir de la glycine (H<sub>3</sub>NCH<sub>2</sub>-) et de la sérine (-CH<sub>2</sub>OH). Différentes structures, qui diffèrent par la position des molécules d'eau sur les sites accepteurs et donneurs de liaison hydrogène, ont été utilisées comme point de départ des dynamiques moléculaires. Pour une même structure de départ, nous avons lancé en parallèle plusieurs dynamiques à des températures différentes (150 et 250 K) avec des temps de dynamique de 0.0016, 0.016 et 0.04 ns pour lesquels les structures ont été sauvegardées tous les 0.016, 0.16 et 0.4 ps, respectivement. Toutes les structures générées lors de ces dynamiques moléculaires ont été optimisées avec AMOEBA.

82 et 306 minimas ont été obtenus pour  $\text{HOCH}_2\text{NH}_2(\text{H}_2\text{O})_n$ , avec  $n = 3$  et 4, respectivement. Pour  $\text{HOCH}_2\text{NH}_3^+(\text{H}_2\text{O})_n$ , 102 et 216 minimas ont été identifiés pour  $n = 3$  et 4. Toutes les structures AMOEBA ont été ensuite ré-optimisées au niveau de théorie B3LYP/6-31G(d). Les fréquences de vibrations, calculées à ce même niveau de théorie, nous ont permis d'obtenir les enthalpies corrigées de la ZPE,  $H_0^\circ$  pour les structures les plus stables situées dans les fenêtres (i) 0-10  $\text{kJ mol}^{-1}$  pour  $\text{HOCH}_2\text{NH}_2(\text{H}_2\text{O})_n$ , avec  $n = 3$  et 4, (ii) 0-15 et 0-25  $\text{kJ mol}^{-1}$  pour  $\text{HOCH}_2\text{NH}_3^+(\text{H}_2\text{O})_3$  et  $\text{HOCH}_2\text{NH}_3^+(\text{H}_2\text{O})_4$  (énergies au niveau B3LYP/6-31+G(d,p)). Ensuite, pour nous assurer que tous les minimas de basses énergies avaient été identifiés, nous avons construit un arbre de Darwin, pour  $n$  allant de 0 à 4. Aucun nouveau minima n'a été généré. Les énergies de ces mêmes structures ont été améliorées avec le niveau de théorie G3B3.

## Profils en enthalpie à 0 K

Les profils en enthalpie à 0 K ont été établis à partir des conformations les plus stables de l'aminomethanol diversement hydraté ( $n = 0$  à 4). Ces structures ont servi de points de départ des recherches des états de transition qui conduisent aux réactifs et aux produits au niveau B3LYP/6-31G(d). La connectivité de chaque état de transition a été contrôlée systématiquement par un calcul IRC. Les énergies des états stables et de transition ont ensuite été améliorées par des calculs au niveau G3B3. Pour les profils les plus intéressants, nous avons ré-optimisé les géométries dans un modèle de solvant continu en utilisant une cavité UFF au niveau de théorie G3B3 (géométrie optimisée en B3LYP/6-31+G(d)). Idéalement, afin de considérer l'effet d'un environnement de glace, la nature du système cristallin et la constante diélectrique qui lui est associée sont nécessaires au calcul. Il est toutefois hautement probable que la structure des glaces "sales" interstellaires est de type amorphe i.e. n'appartient pas à un système cristallin bien déterminé.[34] Son traitement théorique est donc délicat. Fort heureusement les effets sur un calcul C-PCM semblent limités. En effet nous avons effectué quelques calculs en utilisant la constante diélectrique de la glace XI (forme la plus stable de la glace à basse température) c'est-à-dire  $\epsilon = 107$ . [35] Or, nous n'avons pas observé de différence notable sur les structures et sur les énergies critiques (moins de 1  $\text{kJ mol}^{-1}$ ) par rapport à l'eau liquide (constante diélectrique  $\epsilon = 78.4$ ). De ce fait, nous avons conservé la valeur de la constante diélectrique de l'eau liquide pour l'ensemble des calculs.

## V.2.4. Résultats et discussion

Les résultats de notre étude seront présentés sous deux grandes rubriques. Dans la première, les systèmes globalement neutres  $(\text{CH}_2\text{O} + \text{NH}_3) + n\text{H}_2\text{O} \rightarrow (\text{HNCH}_2 + \text{H}_2\text{O}) + n\text{H}_2\text{O}$  (notés **NW0**, **NW1**, **NW2**, **NW3** et **NW4**, pour  $n = 0-4$ ) seront présentés. La voie de dissociation concurrente  $\text{H}_2\text{CNH}_2^+(\text{H}_2\text{O})_{n-m} + \text{OH}^-(\text{H}_2\text{O})_m$  sera également examinée. La seconde partie sera dédiée aux systèmes protonés  $(\text{CH}_2\text{O} + \text{NH}_4^+) + n\text{H}_2\text{O} \rightarrow (\text{H}_2\text{NCH}_2^+ + \text{H}_2\text{O}) + n\text{H}_2\text{O}$  (notés **PW0**, **PW1**, **PW2**, **PW3** et **PW4**, pour  $n = 0-4$ ). De plus, pour ces différents systèmes, une simulation de l'environnement de l'eau liquide ou sous forme de glace sera réalisée au moyen du modèle de continuum C-PCM.

Une molécule d'eau peut faciliter une réaction par un effet catalytique *passif* ou *actif*. La Figure V.2.2 illustre ces effets dans le cas de la formation de l'aminomethanol  $\text{NH}_2\text{CH}_2\text{OH}$  en présence d'une molécule d'eau (système **NW1**). Trois types de catalyses passives peuvent être a priori définis selon que la molécule d'eau joue le rôle de donneur (*acide*) ou d'accepteur (*base*) de liaisons hydrogènes. La molécule d'eau peut aussi "assister" la migration du proton entre l'oxygène et l'azote (catalyse par *transport*). Enfin, la molécule d'eau peut directement participer à la réaction durant laquelle des liaisons O-H et N-H sont créés et(ou) rompues (catalyse *active*). Dans l'ensemble des systèmes étudiés ici, nous avons constaté que trois catégories de processus pouvaient se produire durant lesquels 0, 1 ou 2 molécules d'eau jouent le rôle de catalyse *active* (les autres molécules d'eau ayant alors un rôle *passif*). Nous distinguerons ces trois types de catalyses actives par les sigles **A0**, **A1** et **A2**.

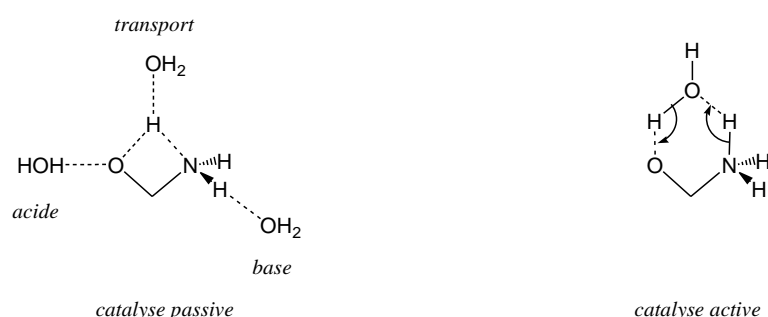


Figure V.2.2. Type de catalyse pour une molécule d'eau.

### ***V.2.4.1. Les premières étapes de la synthèse de Strecker dans un milieu neutre***

#### **Le système non solvaté (NW0)**

Une molécule d'ammoniac peut interagir avec une molécule de formaldéhyde par la formation de liaisons hydrogènes  $\text{NH}\cdots\text{OC}$  ou  $\text{CH}_2\cdots\text{NH}_3$  ou bien par la création d'une liaison covalente dans une forme zwitterionique  $\text{H}_3\text{N}^+-\text{CH}_2\text{O}^-$ . La structure la plus stable, **0A1**, correspond à cette dernière situation, toutefois, la distance  $\text{C}\cdots\text{N}$  est égale à 2.764 Å, il ne peut donc pas s'agir d'une liaison covalente. Une analyse NCI démontre en effet l'existence d'une interaction électrostatique non covalente entre N et C (surface verte, Figure V.2.3). Cette interaction est relativement modeste puisque le complexe **0A1** n'est stabilisé que de 7 kJ mol<sup>-1</sup> par rapport à ses constituants isolés  $\text{NH}_3$  et  $\text{H}_2\text{CO}$ . Partant de **0A1**, la migration 1,3 de proton (étape (a)) conduit à l'aminométhanol dans sa conformation **0B1**. Ce transfert est associé à une barrière énergétique de 140 kJ mol<sup>-1</sup>. L'étape (b) de la réaction de Strecker implique une nouvelle migration 1,3 de proton de N vers OH. Cette réaction n'est possible que si le groupement  $\text{NH}_2$  de l'aminométhanol est dans une position adéquate par rapport à la liaison CO. Une rotation du groupe amino permet de passer de la conformation **0B1** à **0B2**. L'état de transition de cette rotation est caractérisé par une barrière énergétique de 16 kJ mol<sup>-1</sup>. La formation du méthylèneimine à partir de l'aminométhanol dans sa conformation **0B2** (étape (b)) nécessite le franchissement d'une barrière énergétique importante de 236 kJ mol<sup>-1</sup>. L'état de transition correspondant, **0b0**, conduit au complexe électrostatique **0C1** qui doit sa stabilité, par rapport à ses constituants, à l'existence d'une liaison hydrogène,  $\text{HOH}\cdots\text{N(H)=CH}_2$ .

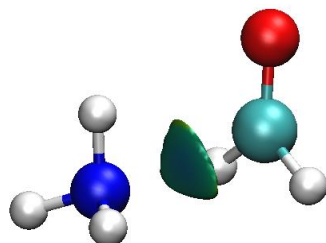


Figure V.2.3. Illustration des interactions non-covalentes au sein de **0A1**. La densité SCF calculée au niveau B3LYP/6-31G(d) a été utilisée et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.

Le profil en enthalpie à 0 K des étapes (a) et (b) de la réaction "sèche"  $\text{CH}_2\text{O} + \text{NH}_3 \rightarrow \text{CH}_2\text{NH} + \text{H}_2\text{O}$  est résumé sur la Figure V.2.4. Les enthalpies relatives à 0 K de **0a0** et **0b0** par rapport aux réactifs **0R** sont égales à 133 et 199 kJ mol<sup>-1</sup> au niveau G3B3. Ces valeurs sont proches de celles obtenues dans la littérature au meilleur niveau de théorie (CCSD(T)/cc-pVDZ, 128 et 190 kJ mol<sup>-1</sup> respectivement [30]). On notera que, compte tenu des hautes barrières associées aux deux migrations 1,3 de protons, **0B** est une espèce thermodynamiquement stable qui devrait donc être détectable dans le milieu interstellaire.

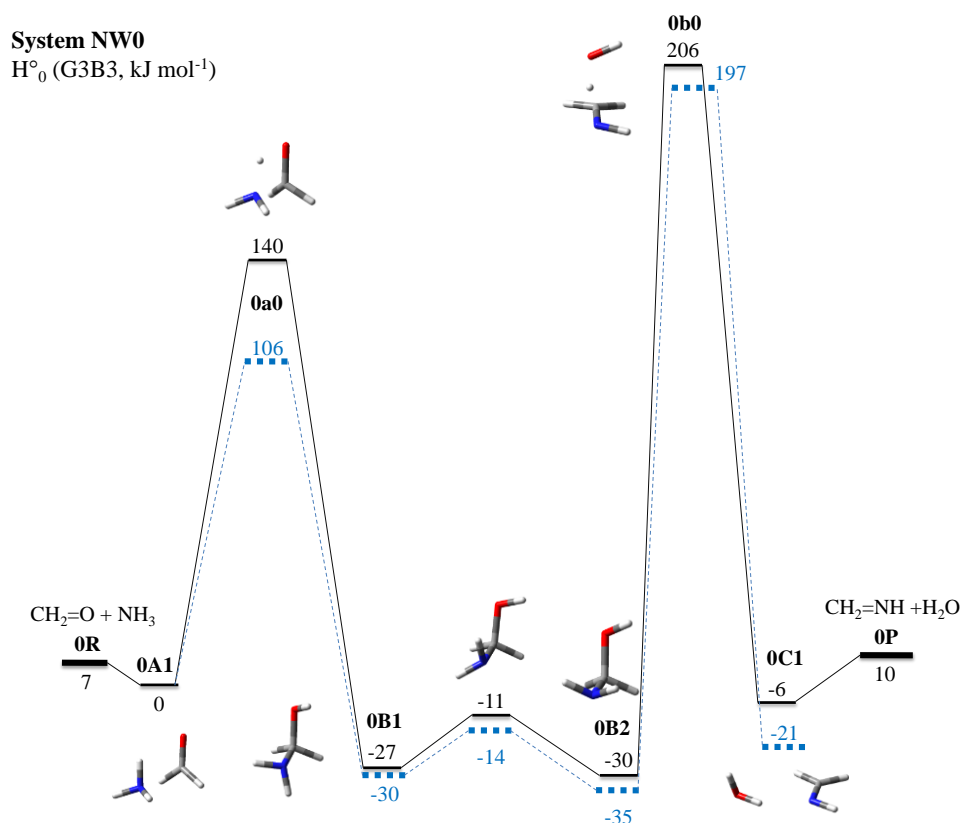


Figure V.2.4. Profil en enthalpie à 0 K en G3B3 pour le système NW0. Les lignes en pointillés correspondent aux calculs C-PCM.

La prise en compte globale du solvant (eau liquide ou glace) avec le modèle C-CPM/UFF a pour effet de diminuer les distances N<sup>⋯</sup>C et OH<sup>⋯</sup>N de **0A1** et **0C1** de 0.085 Å. De plus, la solvation induit une augmentation de l'angle OH<sup>⋯</sup>N dans **0C1** (de 149° à 177°). Enfin, dans la structure des états de transition, les longueurs N<sup>⋯</sup>H<sup>⋯</sup>O sont légèrement modifiées (de 0.020 Å). Seule l'énergie critique de la première étape (a) est influencée par le continuum. L'état de transition correspondant est abaissé de ~ 30 kJ mol<sup>-1</sup> (Figure V.2.4).



## Le système mono-hydraté (NW1)

Nous avons déterminé le profil en enthalpie à 0 K de la formation du méthylèneimine à partir du formaldéhyde mono-hydraté (voir Figure V.2.4) pour les deux types de catalyses possibles : **A1**, la molécule d'eau est "active" et **A0**, la molécule d'eau est "passive". Pour cette dernière approche, nous avons testé les trois possibilités de catalyse passive (acide, basique et transport voir Figure V.2.2), mais toutes nos tentatives ont convergé vers les mêmes espèces où la molécule d'eau interagit avec les groupements NH et CO. Comme attendu, la catalyse passive **A0** est moins efficace que la catalyse active **A1**. En effet, les états de transition correspondants, **1a0** et **1b0** sont situés respectivement 56 et 64 kJ mol<sup>-1</sup> au-dessus des états de transition du trajet "catalyse active", **1a1** et **1b1** (Figure V.2.5). On remarquera toutefois que les enthalpies critiques de la catalyse passive (119 et 179 kJ mol<sup>-1</sup> trajet **A0**) sont un peu plus basses que celles calculées pour le système non-hydraté (140 et 206 kJ mol<sup>-1</sup>, Figure V.2.4), démontrant ainsi son efficacité même si elle est limitée.

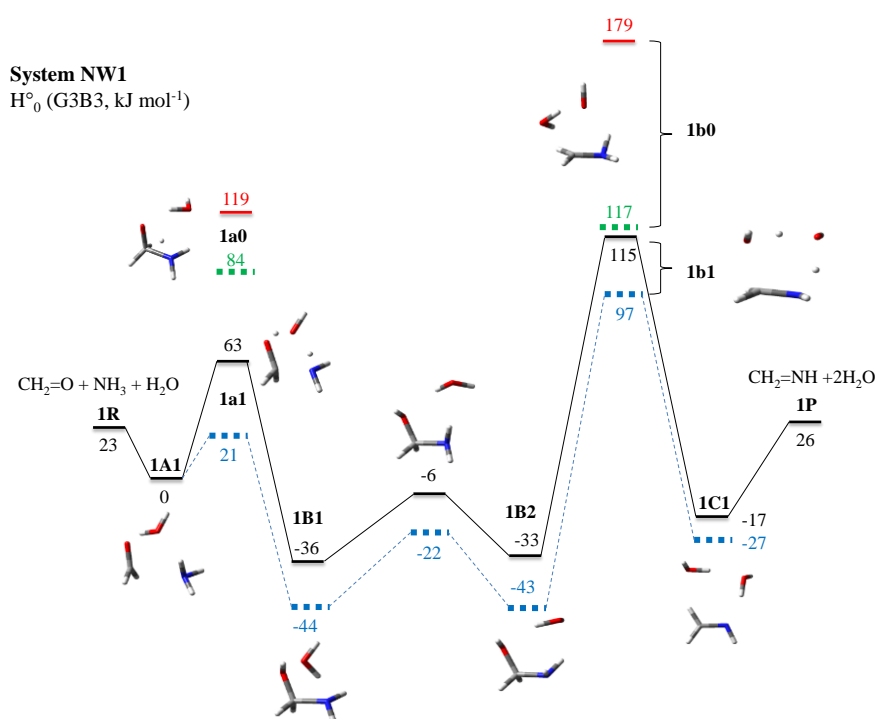


Figure V.2.5. Profil en enthalpie à 0 K en G3B3 pour le système NW1. (i) **A1** : « catalyse active » par 1 molécule de H<sub>2</sub>O (**1a1** et **1b1**), et (ii), **A0** : aucune catalyse active mais 1 molécule de H<sub>2</sub>O spectateur (**1a0** et **1b0**). Les lignes pointillées correspondent aux calculs C-PCM sur le système **A1**.

D'un point de vue mécanistique, le processus global de catalyse active **A1** du système NW1 fait intervenir quatre complexes comme indiqué sur la Figure V.2.6.

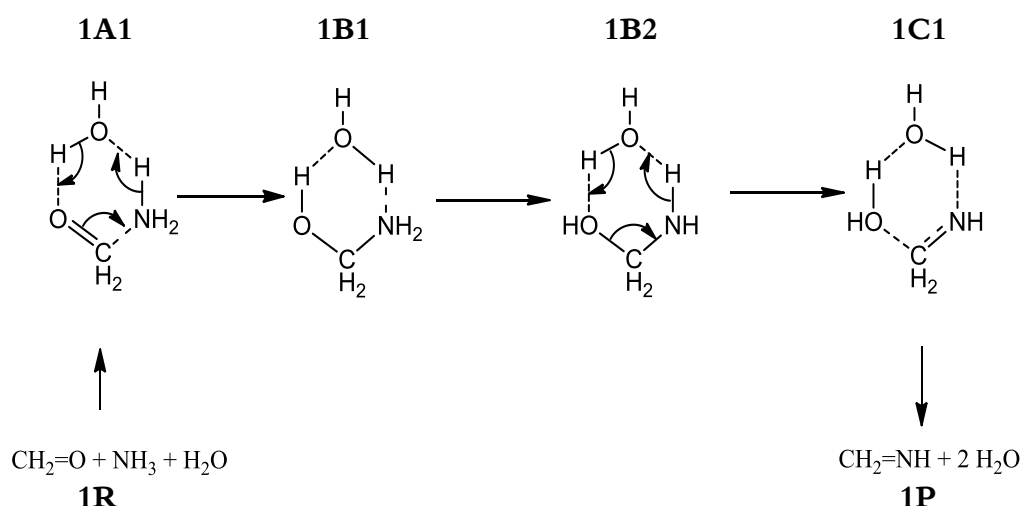


Figure V.2.6. Résumé des mécanismes de transfert de proton au cours de la formation du méthylèneimine par catalyse active (trajet **NW1-A1**).

Le complexe d'approche **1A1** dans lequel la molécule de H<sub>2</sub>O pontre les groupements NH et CO est stabilisé par deux liaisons hydrogènes comme l'indique l'analyse NCI (surface bleue, Figure V.2.7). L'énergie de stabilisation de **1A1** est égale à  $\sim 20 \text{ kJ mol}^{-1}$  par rapport à ses constituants. Par comparaison avec le système **NW0**, la liaison C...N est ici plus courte (2.538 Å dans **1A1** contre 2.764 Å dans **0A1**), l'interaction non-covalente y est plus importante. Les conformations **1B1** et **1B2** ont des énergies similaires et sont notablement plus stables que **1C1** et **1A1**. Cette situation est comparable à ce qui avait été observé pour le système **NW0** avec toutefois un léger gain de stabilité ( $\sim 10 \text{ kJ mol}^{-1}$ ) du fait de la présence de la molécule d'eau. En revanche, les enthalpies à 0 K de **1a1** et **1b1** sont fortement abaissées en présence de la molécule d'eau active. Cela se traduit par une diminution des énergies critiques à 0 K correspondantes de 77 et 91 kJ mol<sup>-1</sup>. L'effet catalytique de la molécule d'eau excédentaire dans le système **NW1** est probablement lié au caractère coopératif des liaisons hydrogènes dans les diverses structures **1A1-1C1** et à la mise en œuvre de transferts de protons concertés dans une configuration particulièrement favorable à six centres.

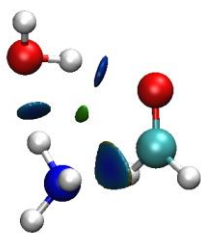


Figure V.2.7. Illustration des interactions non-covalentes au sein de **1A1**. La densité SCF calculée au niveau B3LYP/6-31G(d) a été utilisée et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.

Comme nous l'avons observé dans le profil associé à **NW0**, la prise en compte de l'environnement par C-PCM modifie principalement la barrière de l'étape (a) (**1A1**→**1a1** passe de 63 à 21 kJ mol<sup>-1</sup>), tandis que celle associée à l'étape (b) reste pratiquement inchangée (**1B2**→**1b1** passe de 148 à 140 kJ mol<sup>-1</sup>).

La diminution spectaculaire des barrières correspondant aux étapes (a) et (b) lorsqu'on passe du système **NW0** à **NW1** n'est pas suffisante pour que la réaction de Strecker se fasse spontanément même en présence d'une catalyse active par une molécule d'eau. L'étape déterminante correspond encore à l'étape (b).

## Le système di-hydraté (NW2)

Les combinaisons impliquant le formaldéhyde, l'ammoniac et deux molécules d'eau sont nombreuses. Nous ne présenterons que les complexes **2A1** et **2A2**, précurseurs des deux trajets de plus basse énergie. Ces deux complexes se différencient par la disposition "en parallèle" (**2A1**) ou "en série" (**2A2**) des deux molécules d'eau qui interagissent avec le système résiduel CH<sub>2</sub>O<sup>···</sup>NH<sub>3</sub>. Il est intéressant de noter que la distance C<sup>···</sup>N est égale à 1.682 Å dans **2A1** et à 2.367 Å dans **2A2**. La structure **2A1** se rapproche donc d'un zwitterion. Cela est confirmé par l'analyse NCI qui démontre qu'il n'existe pas d'interaction à faible portée entre C et N. En revanche, deux interactions de type van der Waals sont observées entre NH et OC, ainsi qu'entre OH et NH (surface verte, Figure V.2.8). Dans le complexe **5A4**, les deux molécules d'eau et les réactifs CH<sub>2</sub>O et NH<sub>3</sub> forment un système coopératif de liaisons hydrogènes combiné à l'interaction non covalente CH<sub>2</sub>O<sup>···</sup>NH<sub>3</sub> (surfaces bleues). Cette disposition conduit au système le plus stable (c'est le point de référence en H°<sub>0</sub> sur les Figure V.2.9 et V.2.10).

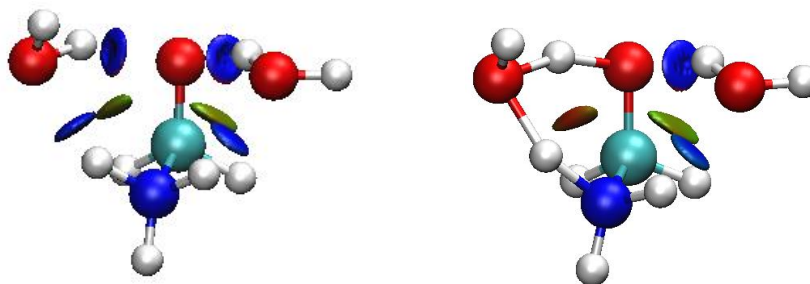


Figure V.2.8. Illustration des interactions non-covalentes au sein du complexe **2A1** et **2a1**. La densité SCF calculée au niveau B3LYP/6-31G(d) a été utilisée et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.

Trois possibilités de catalyses, impliquant zéro, une ou deux molécules d'eau actives (**A0**, **A1** et **A2**), ont été envisagées, les profils en enthalpie à 0 K sont présentés sur les Figures V.2.9 et V.2.10.

*trajet NW2-A0.* L'absence de molécules d'eau active dans le processus réactionnel est ici encore très défavorable. Nous n'avons pas réussi à localiser un état de transition pour l'étape (a). L'état de transition identifié pour l'étape (b), **2b0**, est situé  $162 \text{ kJ mol}^{-1}$  au-dessus des réactifs isolés ce qui le place plus de  $100 \text{ kJ mol}^{-1}$  au-dessus des états de transitions des voies **A1** et **A2** discutées ci-dessous (Figure V.2.9).

*trajet NW2-A1.* Le trajet complet permettant de passer des réactifs aux produits de la réaction de Strecker en présence d'une molécule d'eau active et d'une molécule d'eau passive a pu être établi. Les étapes (a) et (b) sont associées aux états de transition **2a1** et **2b1** situés  $5 \text{ kJ mol}^{-1}$  en dessous et  $61 \text{ kJ mol}^{-1}$  au-dessus des réactifs isolés, respectivement. Une comparaison entre les systèmes **NW1-A1** et **NW2-A1** est instructive. En effet, la barrière à franchir pour l'étape (b) est pratiquement identique pour les systèmes **NW1** et **NW2**, ( $148$  et  $144 \text{ kJ mol}^{-1}$ ), tandis que l'étape (a) est nettement favorisée par l'ajout d'une molécule d'eau passive (l'énergie critique passe de  $63$  à  $16 \text{ kJ mol}^{-1}$ ) puisque par sa présence la distance C...N est plus courte ( $1.682 \text{ \AA}$ ) que dans le complexe **1A1** ( $2.538 \text{ \AA}$ ) facilitant ainsi la formation de la liaison covalente C-N.

*trajet NW2-A2.* Par rapport au trajet **NW2-A1**, la participation active d'une deuxième molécule d'eau se traduit par un abaissement en enthalpie des états de transition des deux

étapes (a) et (b). Cet effet est surtout notable pour l'étape (b) où l'abaissement atteint 16 kJ mol<sup>-1</sup>.

La prise en compte de l'effet de l'environnement par la méthode C-PCM a plus d'influence (i) sur les barrières associées à l'étape (a) et (ii) sur les états de transitions du trajet **NW2-A2**. Le gain en enthalpie passe de 21 à 36 kJ mol<sup>-1</sup> lorsqu'on passe de **NW2-A1** à **NW2-A2** pour l'étape (a) et de 12 à 28 kJ mol<sup>-1</sup> pour l'étape (b).

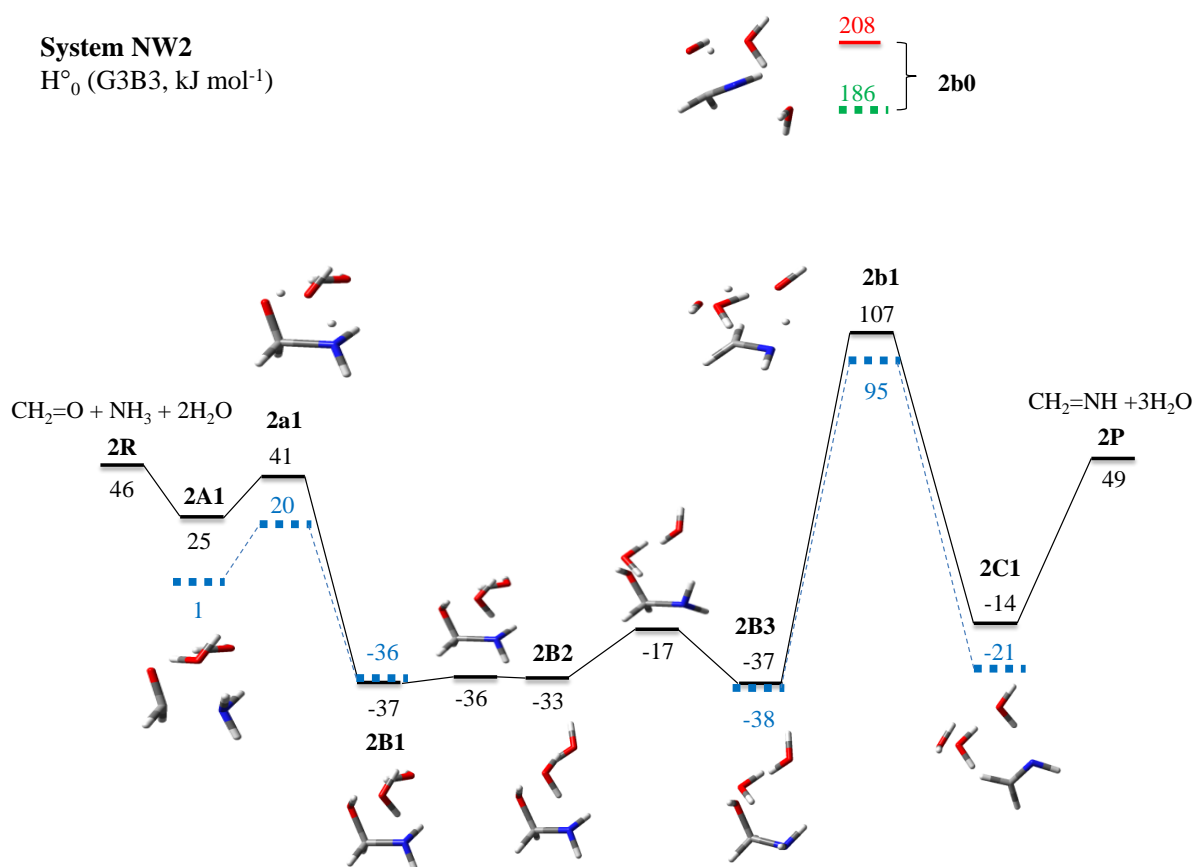


Figure V.2.9. Profil en enthalpie à 0 K en G3B3 pour le système **NW2**. (i) **A1** : « catalyse active » par 1 molécule de H<sub>2</sub>O plus 1 molécule de H<sub>2</sub>O spectateur (**2a1** et **2b1**), et (ii), **A0** : aucune catalyse active, les 2 molécules d'eau jouent un rôle passif (**2b0**). Les lignes en pointillés correspondent aux calculs C-PCM.

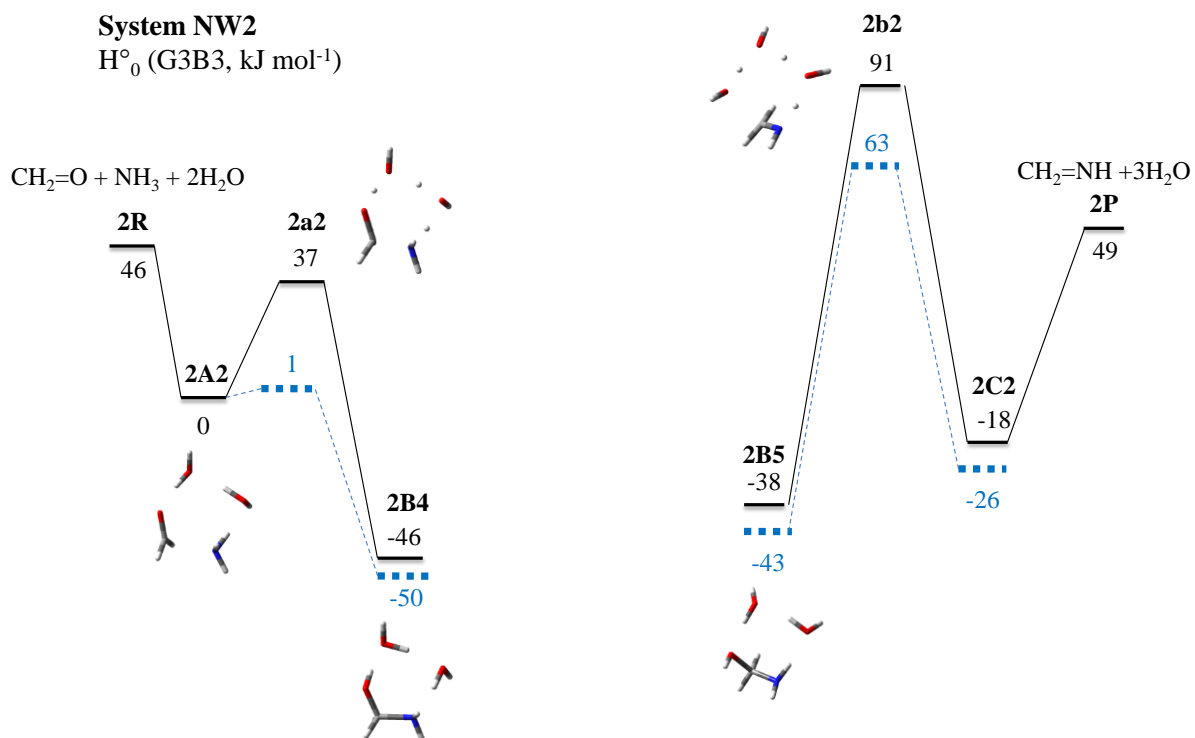


Figure V.2.10. Profil en enthalpie à 0 K en G3B3 pour le système NW2. **A2** : « catalyse active » par 2 molécules de H<sub>2</sub>O. Les lignes pointillées correspondent aux calculs C-PCM.

## Le système tri-hydraté (NW3)

Nous avons identifié plus de 10 structures impliquant CH<sub>2</sub>O, NH<sub>3</sub> et trois molécules d'eau. Parmi celles-ci, le complexe **3A1** est le plus stable. Il s'agit d'un zwitterion et non pas d'un complexe électrostatique (la liaison C<sup>+</sup>...N est égale à 1.636 Å) stabilisé notamment par 6 liaisons hydrogènes avec les trois molécules d'eau (surfaces bleues, Figure V.2.11). La position des molécules de H<sub>2</sub>O dans **3A1** permet d'envisager la réalisation de l'étape (a) par catalyse active avec une (NW3-A1) ou deux molécules d'eau (NW3-A2). Le trajet catalytique de type **A0** ne semble pas réalisable à partir de cette structure.

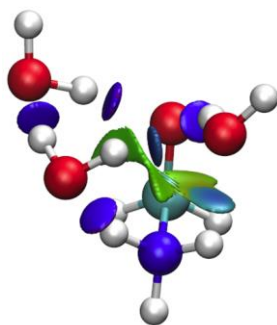


Figure V.2.11. Illustration des interactions non-covalentes au sein du complexe **3A1**. La densité SCF calculée au niveau B3LYP/6-31G(d) a été utilisée et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.

Les profils en enthalpies à 0 K du système **NW3** sont reportés sur les Figures V.2.12 et V.2.13.

*trajet NW3-A0.* L'état de transition de l'étape (**b**), **3b0**, est situé  $180 \text{ kJ mol}^{-1}$  au-dessus du complexe **3A1** soit  $114 \text{ kJ mol}^{-1}$  au-dessus des réactifs isolés. C'est de loin l'état de transition le plus haut du système **NW3**.

*trajet NW3-A1.* Le trajet réactionnel conduisant les réactifs aux produits de la réaction de Strecker en présence d'une molécule d'eau active et de deux molécules d'eau passives est associé aux états de transition **3a1** et **3b1**. Ils sont situés respectivement  $45 \text{ kJ mol}^{-1}$  en dessous et  $5 \text{ kJ mol}^{-1}$  au-dessus des réactifs isolés. Si l'on compare les systèmes **NW2-A1** et **NW3-A1** on constate que la barrière à franchir pour l'étape (**a**) passe de  $16$  à  $21 \text{ kJ mol}^{-1}$  tandis que pour l'étape (**b**) elle passe de  $144$  à  $127 \text{ kJ mol}^{-1}$ . L'incidence de la seconde molécule d'eau spectatrice semble donc plus limitée.

*trajet NW3-A2.* Par rapport au trajet **NW3-A1**, la participation active d'une deuxième molécule d'eau modifie peu la première étape (**a**) mais se traduit par un abaissement en enthalpie d'une quinzaine de  $\text{kJ mol}^{-1}$  pour l'état de transition de l'étape (**b**). Cet effet est surtout notable pour l'étape (**b**) où l'abaissement atteint  $16 \text{ kJ mol}^{-1}$ . Le résultat qui apparaît de façon évidente sur la Figure V.2.13 est que les états de transition des deux étapes deviennent plus bas en enthalpie que les réactifs isolés.





En prenant en compte l'environnement par C-PCM, on constate que les barrières associées aux états de transition de l'étape (a) sont peu modifiées (21 kJ mol<sup>-1</sup> pour **A1** et **A2**). En revanche, les états de transition associés à l'étape (b) sont déstabilisés de 10 à 20 kJ mol<sup>-1</sup> respectivement pour les catalyses **A1** et **A2**.

## Le système tetra-hydraté (NW4)

Comme attendu, à chaque nouvel ajout d'une molécule de H<sub>2</sub>O explicite, l'enthalpie de stabilisation des divers complexes augmente. Dans le cas du système **NW4** le complexe d'approche le plus stable (identifié parmi les 6 structures examinées) est la forme **4A1**. Il s'agit encore d'une forme zwitterionique pour laquelle la distance C...N est égale 1.603 Å (c'est-à-dire identique à celle du complexe **3A1** du système **NW3**). Les catalyses de type **A1** et **A2** ont été examinées, les résultats sont présentés sur les Figures V.2.14 et V.2.15.

*trajet NW4-A1.* Les deux états de transition **4a1** et **4b1** ont des enthalpies inférieures à celle des réactifs isolés (-95 et -22 kJ mol<sup>-1</sup>). Passant du système **NW3-A1** à **NW4-A1**, la molécule d'eau passive supplémentaire entraîne une légère diminution des barrières énergétiques (de l'ordre de 5 kJ mol<sup>-1</sup>).

*trajet NW4-A2.* De la même manière que pour **NW4-A1**, les deux états de transition **4a2** et **4b2** ont des enthalpies inférieures à celle des réactifs isolés (-84 et -54 kJ mol<sup>-1</sup>). La comparaison des systèmes **NW3-A2** à **NW4-A2** montre ici encore une légère diminution des énergies critiques calculées (3 et 10 kJ mol<sup>-1</sup>) pour les deux étapes (a) et (b) à la suite de l'introduction d'une molécule d'eau spectatrice supplémentaire.

Enfin, comme observé lors de l'étude de **NW3**, la prise en compte des effets de l'environnement par C-PCM modifie faiblement les enthalpies des barrières des étapes (a) et (b).

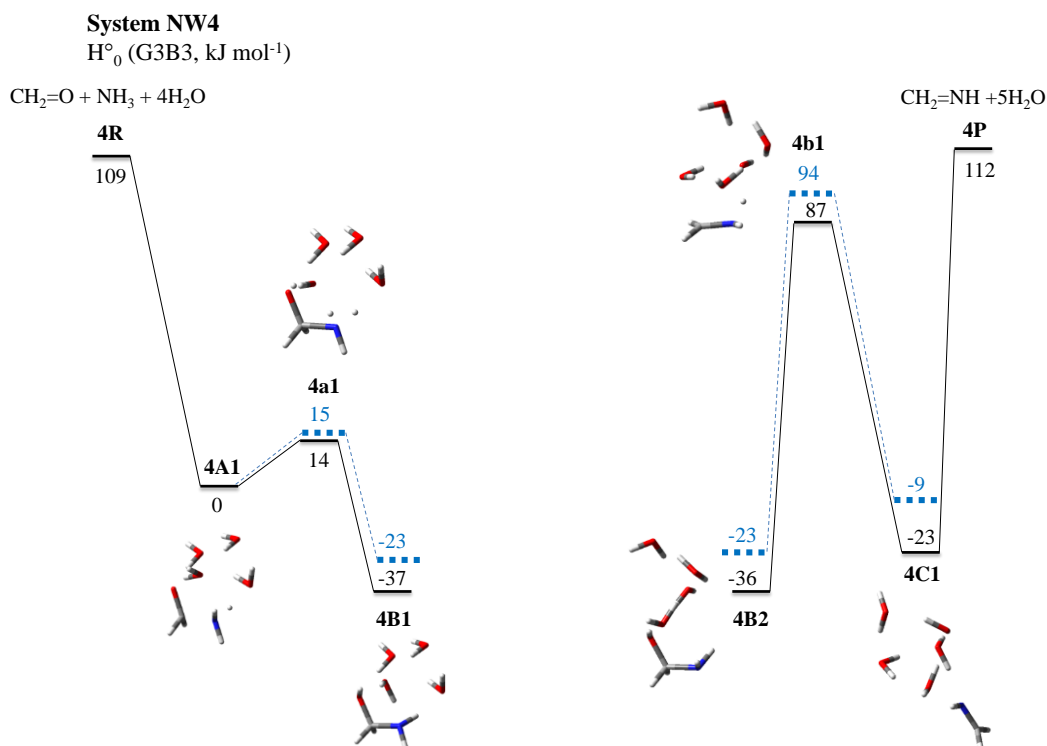


Figure V.2.14. Profil en enthalpie à 0 K en G3B3 pour le système **NW4**. **A1** : « catalyse active » par 1 molécule d'eau plus 3 molécules d'eau spectatrices. Les lignes en pointillés correspondent aux calculs C-PCM.

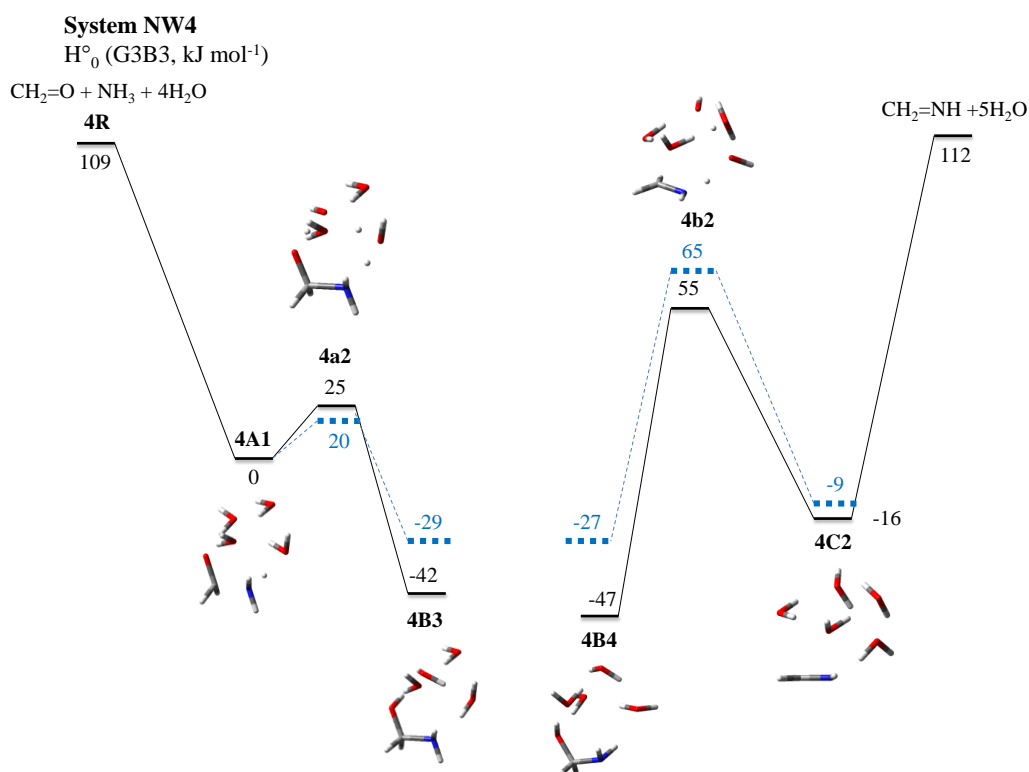


Figure V.2.15. Profil en enthalpie à 0 K en G3B3 pour le système **NW4**. **A2**: « catalyse active » par 2 molécules d'eau plus 2 molécules d'eau spectatrices. Les lignes en pointillés correspondent aux calculs C-PCM.

## Voie de dissociation alternative du type $\text{OH}^- + \text{H}_2\text{CNH}_2^+$ dans un milieu neutre.

Dans un milieu neutre mais polaire, comme l'eau ou la glace, la solvation d'espèces ionisées peut devenir très favorable. Il était donc intéressant de considérer les voies de dissociations des structures intermédiaires  $(\text{H}_2\text{NCCH}_2\text{OH})(\text{H}_2\text{O})_n$  conduisant à la méthylèneimine protonée hydratée  $\text{H}_2\text{CNH}_2^+(\text{H}_2\text{O})_{n-m}$  et aux ions hydroxyles hydratés correspondants  $\text{OH}^-(\text{H}_2\text{O})_m$ .

### *Choix de la cavité*

L'utilisation du modèle C-CPM pour l'étude de systèmes ioniques doit être faite avec la plus grande attention. Le choix de la cavité est déterminant pour le calcul de l'énergie libre de solvation qui correspond à la différence en enthalpie libre de la molécule en solution et en phase gaz ( $\Delta G_{\text{solv}}$ ). [36] Pour l'étude des espèces ioniques, Takano et Houk ont montré que la cavité UAKS est la plus adaptée pour le calcul des énergies libres de solvation des espèces neutres et ionisées à 298 K. Nous avons calculé les  $\Delta G_{\text{solv}}$  à 298 K des ions  $\text{NH}_4^+$  et  $\text{OH}^-$  au niveau G3B3 à l'aide de la méthode C-PCM/UAKS. Nous obtenons  $\Delta G_{\text{solv}}(\text{NH}_4^+) = -341 \text{ kJ mol}^{-1}$  et  $\Delta G_{\text{solv}}(\text{OH}^-) = -431 \text{ kJ mol}^{-1}$ . Ces chiffres sont en assez bon accord avec les  $\Delta G_{\text{solv}}$  expérimentaux de  $-356$ [36] et  $-438$ [37]  $\text{kJ mol}^{-1}$ . Ils présentent toutefois une erreur systématique de l'ordre de  $10 \text{ kJ mol}^{-1}$ . La valeur de  $\Delta G_{\text{solv}}$  que nous calculons pour  $\text{CH}_2\text{NH}_2^+$ ,  $-312 \text{ kJ mol}^{-1}$ , est donc peut être légèrement sous-estimée, en valeur absolue.

### *Applications*

Nous avons reporté dans le Tableau V.2.2 les enthalpies relatives à 0 K des différents produits de dissociation attendus. Les structures des complexes  $\text{OH}^-(\text{H}_2\text{O})_2$ ,  $\text{OH}^-(\text{H}_2\text{O})_3$  et  $\text{OH}^-(\text{H}_2\text{O})_4$  sont représentées dans la Figure V.2.16.

Tableau V.2.2. Enthalpies relatives calculées au niveau G3B3. La référence de chaque **NWn** (n = 0 à 4) est le complexe d'approche A le plus stable.

Systèmes	H° <sub>0</sub> (kJ mol <sup>-1</sup> )	
	Phase gazeuse	C-PCM/UAKS
<b>NW0</b>		
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> + OH <sup>-</sup>	779	60
<i>CH<sub>2</sub>NH + H<sub>2</sub>O</i>	<i>10</i>	<i>-18</i>
<b>NW1</b>		
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) + OH <sup>-</sup>	751	59
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> + OH <sup>-</sup> (H <sub>2</sub> O)	681	80
<i>CH<sub>2</sub>NH + 2H<sub>2</sub>O</i>	<i>26</i>	<i>-31</i>
<b>NW2</b>		
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>2</sub> + OH <sup>-</sup>	739	98
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) + OH <sup>-</sup> (H <sub>2</sub> O)	661	126
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> + OH <sup>-</sup> (H <sub>2</sub> O) <sub>2</sub>	630	77
<i>CH<sub>2</sub>NH + 3H<sub>2</sub>O</i>	<i>49</i>	<i>3</i>
<b>NW3</b>		
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>3</sub> + OH <sup>-</sup>	666	71
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>2</sub> + OH <sup>-</sup> (H <sub>2</sub> O)	645	125
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) + OH <sup>-</sup> (H <sub>2</sub> O) <sub>2</sub>	606	83
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> + OH <sup>-</sup> (H <sub>2</sub> O) <sub>3</sub>	586	76
<i>CH<sub>2</sub>NH + 4H<sub>2</sub>O</i>	<i>69</i>	<i>-2</i>
<b>NW4</b>		
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>4</sub> + OH <sup>-</sup>	670	91
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>3</sub> + OH <sup>-</sup> (H <sub>2</sub> O)	597	129
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) <sub>2</sub> + OH <sup>-</sup> (H <sub>2</sub> O) <sub>2</sub>	613	114
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> (H <sub>2</sub> O) + OH <sup>-</sup> (H <sub>2</sub> O) <sub>3</sub>	586	112
CH <sub>2</sub> NH <sub>2</sub> <sup>+</sup> + OH <sup>-</sup> (H <sub>2</sub> O) <sub>4</sub>	573	109
<i>CH<sub>2</sub>NH + 5H<sub>2</sub>O</i>	<i>112</i>	<i>23</i>

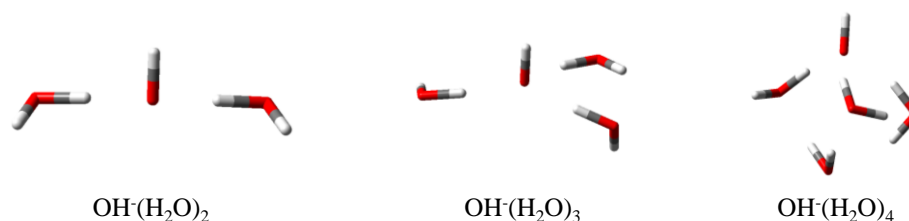


Figure V.2.16. Géométries B3LYP/6-31G(d) des clusters OH<sup>-</sup>(H<sub>2</sub>O)<sub>2</sub>, OH<sup>-</sup>(H<sub>2</sub>O)<sub>3</sub> et OH<sup>-</sup>(H<sub>2</sub>O)<sub>4</sub> en phase gazeuse.

Première observation, en phase gazeuse, la complexation de l'ion OH<sup>-</sup> par une molécule d'eau est considérablement plus stabilisante que la complexation de l'ion iminium CH<sub>2</sub>NH<sub>2</sub><sup>+</sup> (-113 et

-44 kJ mol<sup>-1</sup> à 0 K, soit une différence de 70 kJ mol<sup>-1</sup>). Il n'est donc pas étonnant de constater que les produits CH<sub>2</sub>NH<sub>2</sub><sup>+</sup>(H<sub>2</sub>O)<sub>n-m</sub> + OH<sup>-</sup>(H<sub>2</sub>O)<sub>m</sub> les plus stables correspondent aux plus grandes valeurs de *m*. Avec le solvant, la tendance est similaire puisque ΔH<sup>°</sup><sub>solv</sub>(OH<sup>-</sup>) = -431 kJ mol<sup>-1</sup> et ΔH<sup>°</sup><sub>solv</sub>(CH<sub>2</sub>NH<sub>2</sub><sup>+</sup>) = -310 kJ mol<sup>-1</sup>) et qu'il est donc plus bénéfique de solvater OH<sup>-</sup> que CH<sub>2</sub>NH<sub>2</sub><sup>+</sup>. Il en est de même des ions complexés par une à quatre molécules d'eau. Ceci étant, il est clair à l'examen du Tableau V.2.2 que les produits ionisés CH<sub>2</sub>NH<sub>2</sub><sup>+</sup>(H<sub>2</sub>O)<sub>n-m</sub> + OH<sup>-</sup>(H<sub>2</sub>O)<sub>m</sub> sont toujours plus haut en enthalpie que les produits neutres CH<sub>2</sub>NH(H<sub>2</sub>O)<sub>n-m</sub> + (H<sub>2</sub>O)<sub>m</sub>. Même si elles sont réduites en phase condensée par rapport à la phase gazeuse, les différences ne sont pas inférieures à ~70 kJ mol<sup>-1</sup> (~460 en phase gaz). Ainsi la réaction ne devrait pas évoluer vers la voie de dissociation H<sub>2</sub>CNH<sub>2</sub><sup>+</sup>(H<sub>2</sub>O)<sub>n-m</sub> + OH<sup>-</sup>(H<sub>2</sub>O)<sub>m</sub> aussi bien en phase gazeuse que dans l'eau ou la glace.

## Bilan

Pour résumer l'essentiel des résultats obtenus sur les systèmes **NWn**, nous avons reporté sur la Figure V.2.17 les enthalpies critiques à 0 K calculées pour les étapes (a) et (b) en fonction du nombre total de molécules d'eau. Il apparaît clairement que l'étape (b) est toujours l'étape énergétiquement limitante. Il est d'autre part évident que la catalyse exclusivement passive, de type **A0**, est défavorable par rapport aux catalyses actives **A1** ou **A2**. Ces deux dernières conduisent à des valeurs d'enthalpies critiques similaires comme cela a été souligné dans les discussions précédentes. Enfin on remarquera que l'ajout de molécules d'eau spectatrices ne se traduit pas par une réduction modérée des enthalpies critiques (elle est même inexistante pour de l'étape (b) dans le cas **A0**).

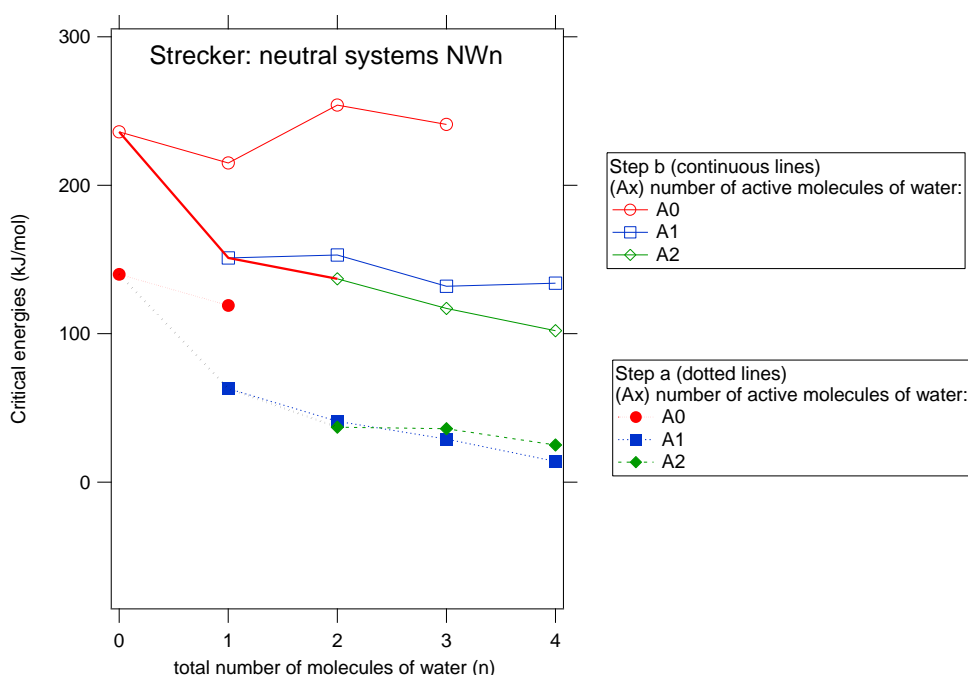


Figure V.2.17. Énergies critiques calculées au niveau G3B3 pour les systèmes neutres **NW<sub>n</sub>** comme une fonction du nombre de molécules d'eau, en phase gazeuse.

La faisabilité de la réaction de Strecker va dépendre de l'état initial des réactifs et de leur éventuelle activation (collisions, absorption de rayonnement électromagnétique, énergie "chimique"). Si l'on considère, comme point de départ, les réactifs dans leurs structures complexées les plus stables (i.e. **0A1**, **1A1**, **2A2**, **3A1** et **4A1**), la réaction sera limitée énergétiquement par l'état de transition le plus bas conduisant à l'imine, également complexée. Les barrières globales correspondantes sont alors égales à 206, 115, 91, 56 et 55 kJ mol<sup>-1</sup> pour les systèmes **NW0** à **NW4**. Pour les trois derniers cas, l'étape (**b**) comporte deux molécules d'eau actives (catalyse **A2**) et semble évoluer vers une valeur asymptotique de ~50 kJ mol<sup>-1</sup>. Les mêmes tendances sont observées lorsque l'on considère le continuum du solvant (eau ou glace). On doit donc en conclure que la réaction de Strecker se heurte à une barrière d'une cinquantaine de kJ mol<sup>-1</sup> et que sa réalisation nécessite une activation préalable.

L'aspect cinétique devrait également être pris en considération dans ces divers ensembles de réactions compétitives. Dans le cadre de ce travail de thèse, nous ne l'avons pas envisagé mais il serait intéressant de le traiter dans de futurs travaux.

### V.2.4.3. Les premières étapes de la synthèse de Strecker dans un milieu acide

#### Le système non solvaté (PW0)

Une étude théorique *ab initio* avait été réalisée sur le système  $\text{H}_2\text{CO} + \text{NH}_3$  en présence d'un proton.[38] Plusieurs complexes stables,  $\text{H}_2\text{CO}\cdots\text{H}^+\cdots\text{NH}_3$ ,  $\text{H}_2\text{C}=\text{HN}\cdots\text{H}^+\cdots\text{OH}_2$ ,  $\text{HOCH}_2^+\cdots\text{NH}_3$  et  $\text{H}_2\text{NCH}_2\cdots\text{OH}_2^+$  (énergies relatives 0, 3, 14 et 32  $\text{kJ mol}^{-1}$ ) ont été identifiés. Les complexes  $\text{H}_2\text{CO}\cdots\text{H}^+\cdots\text{NH}_3$ ,  $\text{H}_2\text{C}=\text{HN}\cdots\text{H}^+\cdots\text{OH}_2$ , et  $\text{HOCH}_2^+\cdots\text{NH}_3$  ont été mis en évidence expérimentalement par spectrométrie de masse tandem.[39] Au niveau de calcul G3B3, le complexe  $\text{H}_2\text{CO}\cdots\text{H}^+\cdots\text{NH}_3$ , **0A1**, est stabilisé par 77  $\text{kJ mol}^{-1}$  par rapport aux réactifs isolés les plus stables  $\text{NH}_4^+$  et  $\text{CH}_2\text{O}$ . Partant de cette structure, la formation de l'aminométhanol protoné (étape (a) de la réaction de Strecker) résulte d'un transfert de proton et d'une attaque nucléophile de l'ammoniac vers l'atome de carbone du formaldéhyde protoné  $\text{H}_2\text{COH}^+$ . L'état de transition correspondant, **0a0**, est situé 153  $\text{kJ mol}^{-1}$  au-dessus de **0A1** (Figure V.2.18). L'étape suivante (b) implique exclusivement une migration 1,3 de proton; l'enthalpie d'activation à 0 K correspondante est égale à 131  $\text{kJ mol}^{-1}$  (**0b0**, Figure V.2.18). Cette valeur obtenue au niveau G3B3 est du même ordre de grandeur que celles obtenues avec des niveaux de théories plus bas par divers auteurs: 144  $\text{kJ mol}^{-1}$  (B3LYP/6-31G(d,p), sans ZPE) [30] et 122  $\text{kJ mol}^{-1}$  (MP2/6-31G(d,p)) [40]. La forme protonée sur l'oxygène de l'aminométhanol, **0C1**, n'est pas une structure covalente mais un complexe de type  $\text{H}_2\text{O}\cdots\text{CH}_2=\text{NH}_2^+$  stabilisé par une forte interaction électrostatique entre l'oxygène de la molécule d'eau et le carbone de la méthylimine protonée comme l'illustre la Figure V.2.19. On notera que la distance  $\text{H}_2\text{O}\cdots\text{CH}_2$  (2.275 Å) est trop importante pour correspondre à une liaison de covalence pure.

En examinant la Figure V.2.18, nous concluons que, en phase gazeuse et sans une activation extérieure supérieure à 75  $\text{kJ mol}^{-1}$ , la réaction  $\text{H}_2\text{CO} + \text{NH}_4^+ \rightarrow \text{H}_2\text{CNH}_2^+ + \text{OH}_2$  ne peut avoir lieu. Toutefois, la formation de  $\text{H}_2\text{CNH}_2^+ + \text{OH}_2$  semble réalisable à partir des réactifs  $\text{H}_2\text{COH}^+ + \text{NH}_3$  situés environ 70  $\text{kJ mol}^{-1}$  au-dessus de l'état de transition de l'étape limitante (a). Mais il est connu que les réactions de transfert de proton exothermiques, comme  $\text{H}_2\text{COH}^+ + \text{NH}_3 \rightarrow \text{H}_2\text{CO} + \text{NH}_4^+$  sont des réactions rapides s'effectuant à des vitesses proches de la vitesse de collision. Ceci est confirmé par diverses tentatives d'observation de l'ion  $\text{H}_2\text{CNH}_2^+$

à partir de  $\text{H}_2\text{COH}^+ + \text{NH}_3$  dans un spectromètre de masse à résonance cyclotronique ionique: seul l'ion  $\text{NH}_4^+$  résultant du transfert de proton est observé.[38,40,41]

L'effet de l'environnement C-PCM/UAKS sur le système **PW0** se traduit par une forte stabilisation de l'intermédiaire aminomethanol protoné **0B1** ( $69 \text{ kJ mol}^{-1}$ ) et, de manière moins prononcée, de l'état de transition de la deuxième étape **1b0** ( $38 \text{ kJ mol}^{-1}$ ). Une conséquence est une augmentation d'une trentaine de  $\text{kJ mol}^{-1}$  de la barrière associée au transfert de proton, **0B1**→**0C1**.

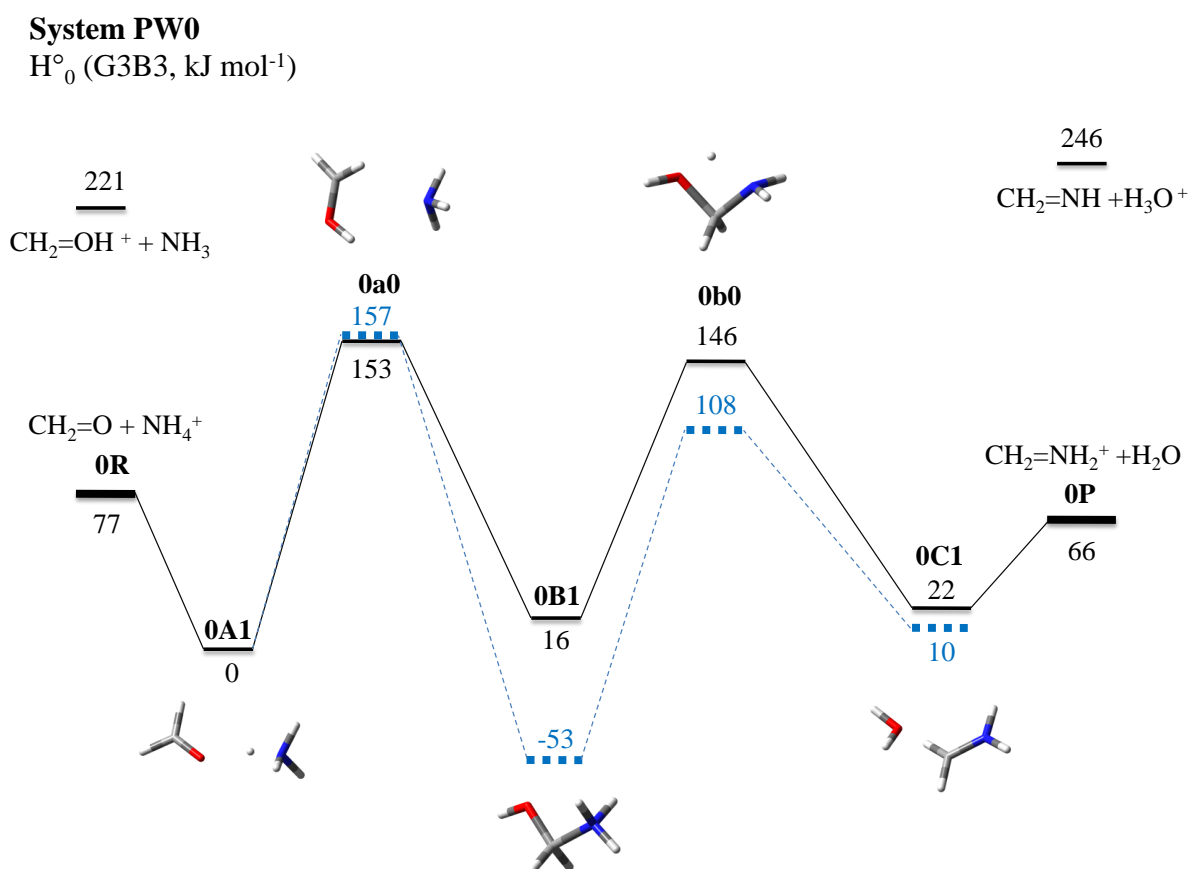


Figure V.2.18. Profil en enthalpie à 0 K en G3B3 pour le système **PW0**. Les lignes en pointillés correspondent aux calculs C-PCM/UAKS.

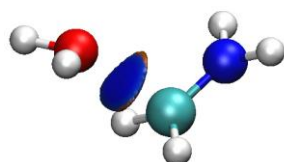


Figure V.2.19. Illustration des interactions non-covalentes au sein de **0C1** (géométrie B3LYP/6-31G(d)). La densité SCF calculée au niveau B3LYP/6-31G(d) a été utilisée et l'échelle de couleur est  $-0.03 < \text{sign}(\lambda_2)\rho < 0.03$  au.



## Systèmes mono-hydratés (PW1)

L'addition d'une molécule d'eau au système  $\text{H}_2\text{CO} + \text{NH}_4^+$  permet de former différents types de complexes, parmi lesquels les deux structures **1A1** et **1A2** (Figure V.2.20). Le complexe le plus stable, **1A2**, est stabilisé par deux liaisons hydrogène dans lesquelles l'ammonium est donneur de proton à la fois pour  $\text{H}_2\text{CO}$  et  $\text{H}_2\text{O}$ . Dans la seconde structure, **1A1**, située  $14 \text{ kJ mol}^{-1}$  au-dessus de **1A2**, la molécule d'eau est donneur de liaison hydrogène pour la molécule  $\text{H}_2\text{CO}$  et accepteur de liaison hydrogène pour  $\text{NH}_4^+$ .

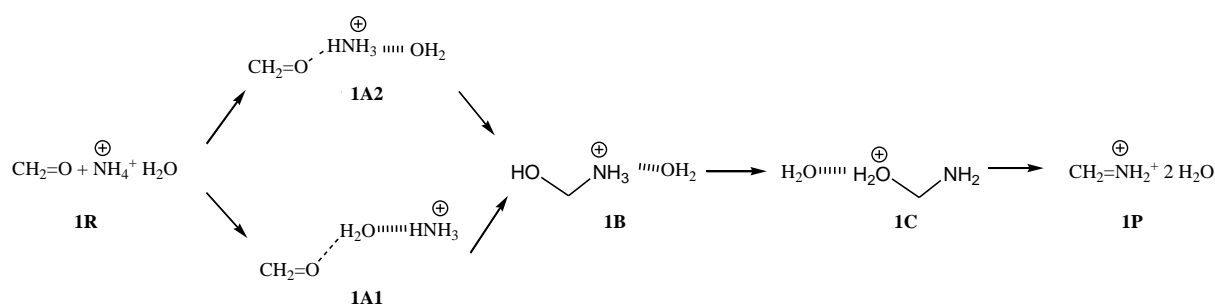


Figure V.2.20. Résumé des mécanismes de transfert de proton au cours de la formation du méthylèneimine protoné par catalyse active (trajet **PW1-A1**).

Le complexe **1A2** est analogue au complexe **0A1** (voir le système **PW0**) solvaté par une molécule d'eau. Cette dernière est éloignée de la zone de transfert du proton et ne peut intervenir que par catalyse passive (**A0**) durant la réaction **1A2**→**1B**. A l'opposé, dans le complexe **1A1**, le transfert de proton peut se faire par le biais de la molécule d'eau située entre  $\text{H}_2\text{CO}$  et  $\text{NH}_4^+$ . Cette voie de catalyse active par une molécule d'eau (**A1**) est la plus favorable d'un point de vue enthalpique (Figure V.2.21). Son état de transition **1a1** est situé  $51 \text{ kJ mol}^{-1}$  en dessous de celui du processus **A0** (**1a0**). De plus, **1a1** est situé  $22 \text{ kJ mol}^{-1}$  en dessous des réactifs isolés  $\text{H}_2\text{CO} + \text{NH}_4^+ + \text{H}_2\text{O}$ .

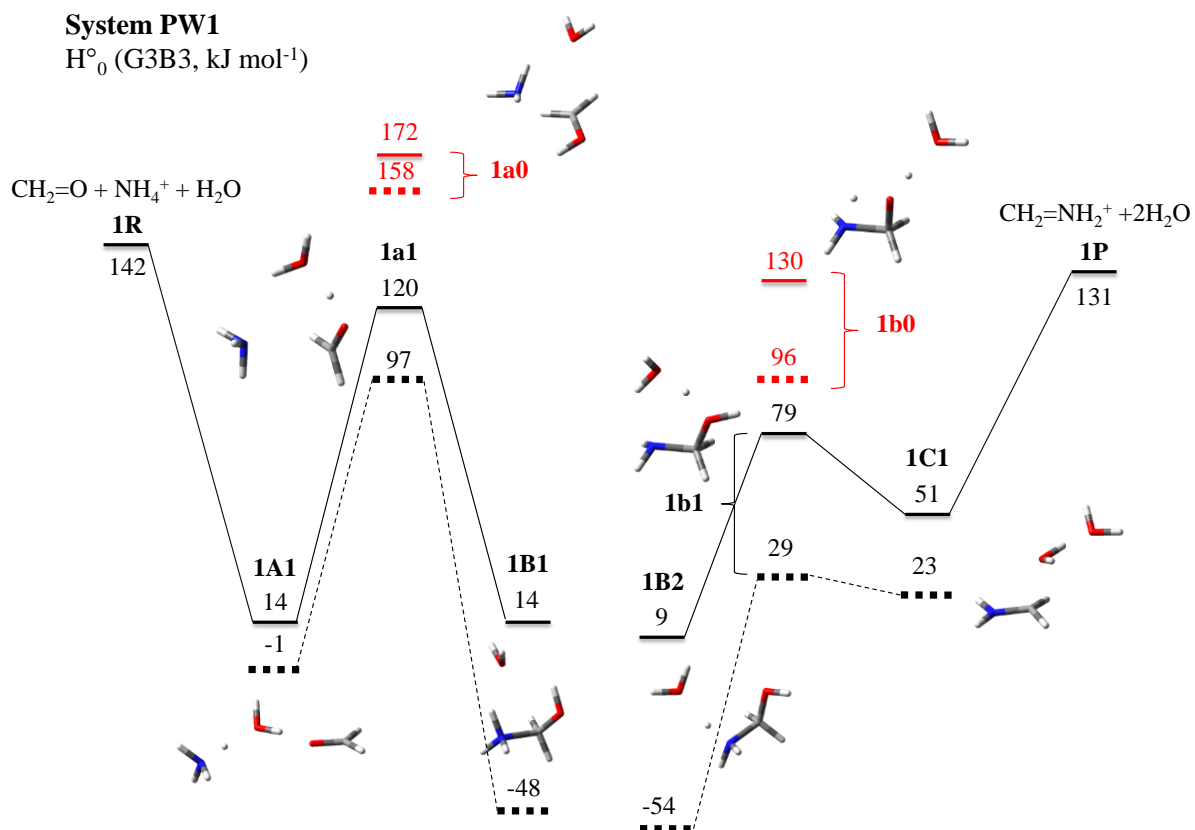


Figure V.2.21. Profil en enthalpie à 0 K en G3B3 pour le système **PW1**. (i) **A1** : « catalyse active » par 1 molécule d'eau (**1a1** et **1b1**), et (ii), **A0** : aucune catalyse active mais 1 molécule d'eau spectatrice (**1a0** et **1b0**). Les lignes en pointillés correspondent aux calculs C-PCM/UAKS.

Une rotation autour de la liaison CO de **1B1** conduit à **1B2** qui par une double migration de proton peut s'isomériser en **1C1**. Lorsque cette réaction (étape (b) de la réaction de Strecker) se déroule selon une catalyse active **A1**, l'état de transition correspondant, **1b1** est situé à 63 kJ mol<sup>-1</sup> en dessous des réactifs isolés. Avec la catalyse passive **A0**, la barrière énergétique est plus élevée (117 kJ mol<sup>-1</sup>) mais l'état de transition correspondant, **1b0** est situé sous les réactifs isolés (mais seulement de 12 kJ mol<sup>-1</sup>).

Si on compare les systèmes **PW0** et **PW1**, on constate qu'une molécule d'eau en catalyse active **A1** diminue de façon conséquente les barrières des étapes (a) et (b) (-47 et -60 kJ mol<sup>-1</sup> respectivement). Autre observation, complémentaire de la précédente, l'ajout de la molécule de H<sub>2</sub>O active stabilise plus fortement l'état de transition de l'étape (b) par rapport aux réactifs isolés. En effet, dans le système **PW0**, **0b0** est plus stable que **0a0** de 7 kJ mol<sup>-1</sup>, tandis que, dans le système **PW1**, **1b1** est plus stable que **1a1** de ~ 40 kJ mol<sup>-1</sup>. Ajoutons enfin que la catalyse passive **A0** a un effet négligeable sur les enthalpies des états de transition correspondants.

La prise en compte globale de l'environnement par C-PCM se traduit ici encore par une forte stabilisation des complexes intermédiaires **B** et des états de transition de l'étape (b) par rapport aux réactifs  $\text{CH}_2\text{O} + \text{NH}_4^+ + \text{H}_2\text{O}$ . Cet effet est beaucoup plus marqué ici que pour le système **PW0**. Cela est observé avec la catalyse passive **A0** (les états de transition des étapes (a) et (b) sont abaissées de 14 et 34  $\text{kJ mol}^{-1}$ , respectivement) et avec la catalyse active **A1** (les états de transition des étapes (a) et (b) sont abaissées de 23 et 50  $\text{kJ mol}^{-1}$ ).

## Systèmes di-hydratés (PW2)

Il existe un grand nombre de complexes résultant de l'association de l'ion ammonium avec le formaldéhyde et deux molécules d'eau. Nous avons choisi de présenter cinq structures pour illustrer les différents cas de figures (**2A1-2A5**, Figure V.2.22).

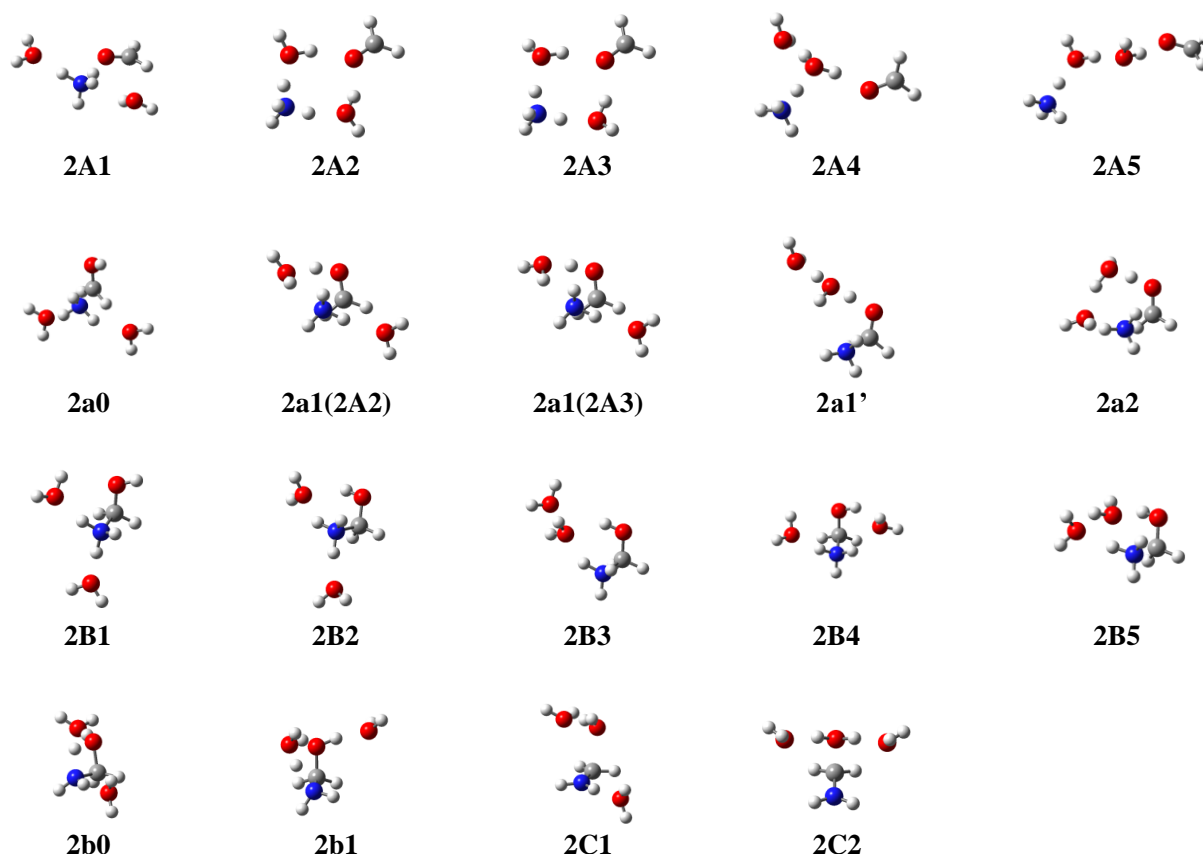


Figure V.2.22. Géométries des structures **PW2** optimisées en B3LYP/6-31G(d) en phase gaz.

Les complexes les plus stables sont les deux formes miroir **2A2** et **2A3**. Ces deux complexes présentent un arrangement cyclique et sont essentiellement stabilisés par quatre liaisons

hydrogène de type  $\text{HOH}\cdots\text{OC}$  et  $\text{NH}\cdots\text{OH}_2$ . **2A4** est un complexe résultant de l'insertion d'une molécule d'eau entre le formaldéhyde et l'ion ammonium, la seconde molécule d'eau interagit avec la première par une interaction de type  $\text{OH}\cdots\text{OH}$ . **2A5** est un complexe d'insertion où  $\text{H}_2\text{CO}$  et  $\text{NH}_4^+$  sont séparés par deux molécules d'eau alignées. **2A1** est une structure analogue à la structure **1A2** envisagée pour le système **PW1**, la deuxième molécule d'eau interagissant avec le formaldéhyde par une forte interaction électrostatique du type  $\text{H}_2\text{O}\cdots\text{CO}$  (distance  $\text{O}\cdots\text{C} = 2.526 \text{ \AA}$ ).

La formation de ces complexes est favorable puisqu'ils sont stabilisés par 165-192  $\text{kJ mol}^{-1}$  par rapport aux réactifs isolés (Figure V.2.23). Cette forte stabilisation entraîne un glissement vers les basses énergies de l'ensemble des intermédiaires des étapes (a) et (b): tous sont situés en dessous des réactifs isolés.

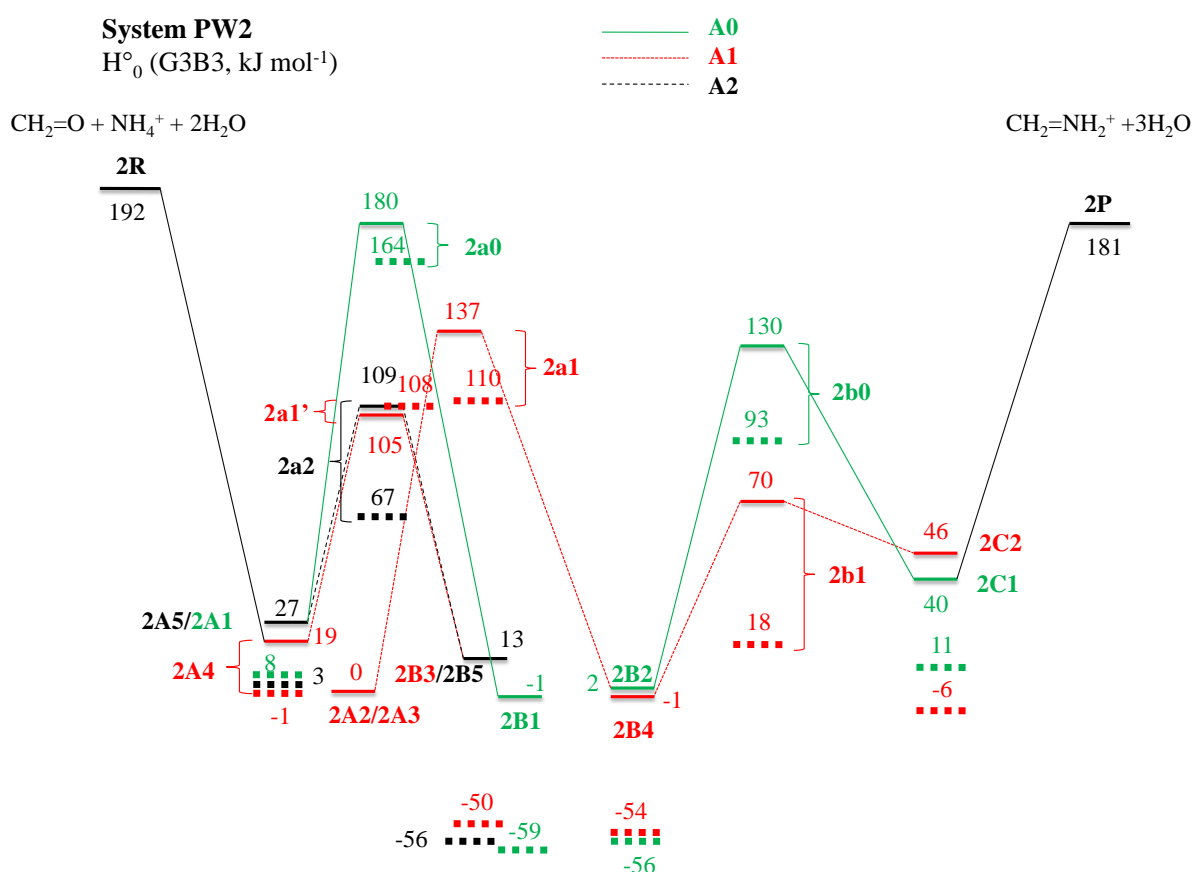


Figure V.2.23. Profil en enthalpie à 0 K en G3B3 pour le système **PW2**. (i) « catalyse active » par 2 molécules d'eau (**A2**, en noir) ; (ii) « catalyse active » par 1 molécule d'eau plus 1 molécule d'eau spectatrice (**A1**, en rouge), et (iii) aucune catalyse active mais 2 molécules d'eau spectatrices (**A0**, en vert). Les lignes en pointillés correspondent aux calculs C-PCM/UAKS.

*trajet PW2-A0.* Les états de transition des étapes (a) et (b), **2a0** et **2b0**, sont les plus hauts en énergie du système **PW2**. Ils sont situés 12 et 62 kJ mol<sup>-1</sup> en dessous des réactifs isolés. On retrouve une différence d'enthalpie proche de 40 kJ mol<sup>-1</sup> entre les deux états de transition observée également pour le système **PW1**, la seconde molécule d'eau passive joue un rôle négligeable.

*trajet PW2-A1.* Deux voies correspondant à l'étape (a) et impliquant une molécule d'eau active et une autre passive ont été identifiées. Les états de transition **2a1'** et **2a1** sont situés respectivement 87 et 55 kJ mol<sup>-1</sup> en dessous des réactifs isolés.

L'état de transition de la seconde étape (b), **2b1**, est encore plus bas puisqu'il est situé 122 kJ mol<sup>-1</sup> sous les réactifs isolés. La disposition des états de transition des étapes (a) et (b) est très proche de celle qui est observée pour le système **PW1-A1** ce qui signifie que l'incidence de la molécule d'eau spectatrice est limitée.

*trajet PW2-A2.* Nous n'avons identifié que l'état de transition de la première étape pour ce type de catalyse: **2a2**. Son enthalpie est pratiquement identique à celle de **2a1'**. Les voies **A1** et **A2** sont donc analogue pour ce système.

## Systèmes tri-hydratés (PW3)

Les complexes d'approche les plus stables du système **PW3**, **3A1** et **3A2**, peuvent être considérés comme étant dérivés du complexe **2A3** du système **PW2** (Figure V.2.24). La troisième molécule d'eau interagit avec NH<sub>4</sub><sup>+</sup> par une interaction de type NH<sup>+</sup>...OH dans **3A1**, tandis que cette même molécule de H<sub>2</sub>O interagit en plus avec une autre molécule de H<sub>2</sub>O par une interaction de type OH...O dans **3A2**. De la même manière, les structures **3B** associées aux mécanismes de moindre énergie sont toutes déductibles de la génération précédente. Ainsi **3B2** résulte de l'ajout d'une molécule de H<sub>2</sub>O sur **2B4**, **3B1** sur **2B4**, **2B2** et **2B1** et **3B3** sur **2B3**.

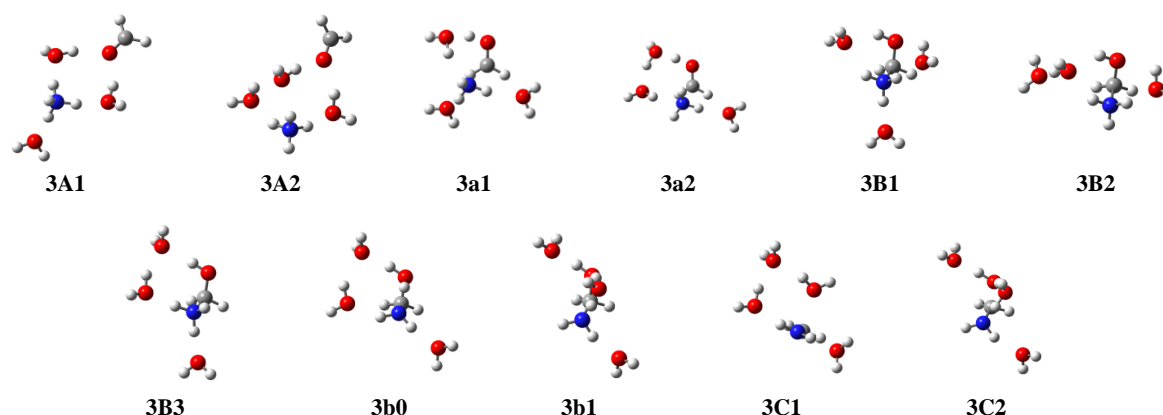


Figure V.2.24. Géométries B3LYP/6-31G(d) des structures **PW3** en phase gazeuse.

La Figure V.2.25 regroupe les résultats obtenus pour le système **PW3**. Un trajet complet, englobant les deux étapes de la réaction de Strecker, a pu être déterminé uniquement pour la catalyse **A1**. Seule l'étape (a) pour **A2** et l'étape (b) pour **A0** ont été identifiées. Néanmoins, toutes les tendances relevées dans les systèmes précédents **PW0**, **PW1** et **PW2** se confirment ici. Ainsi pour un type de catalyse donné, l'état de transition de l'étape (b) est plus bas en enthalpie que celui de l'étape (a). Ensuite, la hauteur des barrières diminue lorsque le nombre de molécules d'eau active augmente. Si on compare les profils énergétiques des systèmes **PW2** et **PW3** (Figures V.2.23 et V.2.25), il apparaît que l'enthalpie relative de l'état de transition de l'étape (a), par rapport aux réactifs isolés, est pratiquement inchangée par l'ajout d'une molécule d'eau spectatrice. En revanche, l'enthalpie relative de l'état de transition de l'étape (b) augmente de façon significative.

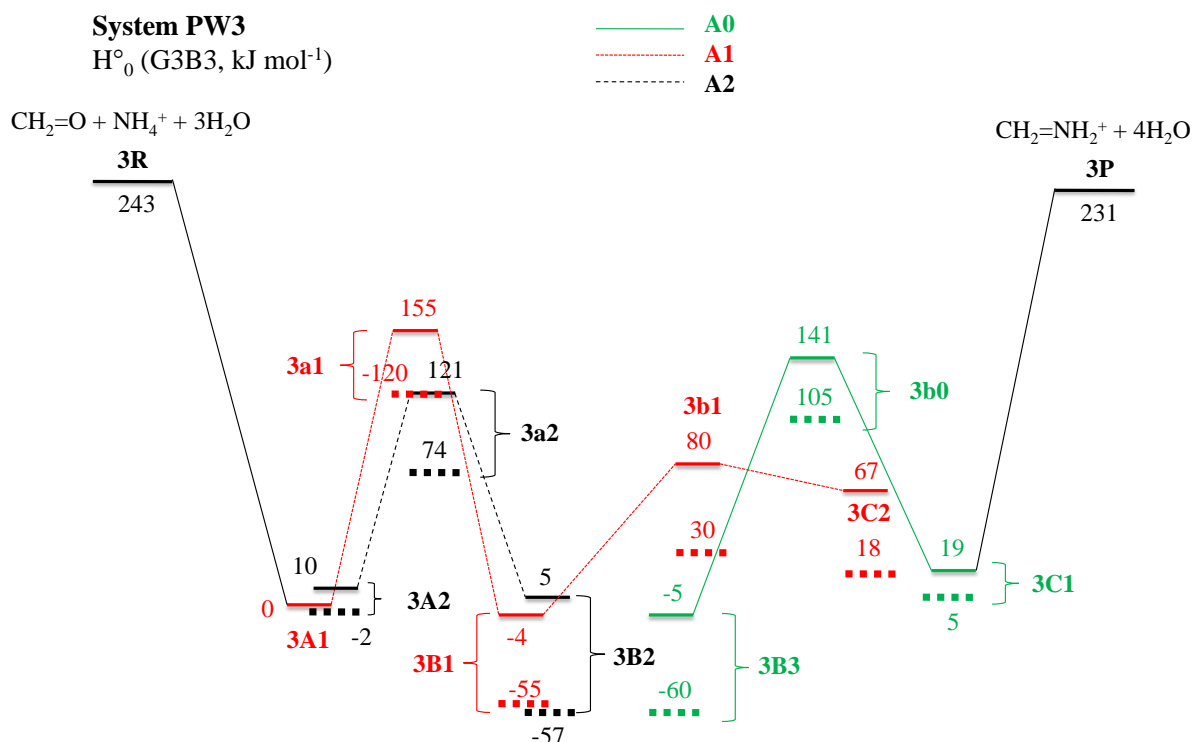


Figure V.2.25. Profil en enthalpie à 0 K en G3B3 pour le système PW3. (i) « catalyse active » par 2 molécules de H<sub>2</sub>O plus 1 molécule de H<sub>2</sub>O spectateur (A2, en noir) ; (ii) « catalyse active » par 1 molécule de H<sub>2</sub>O plus 2 molécule de H<sub>2</sub>O spectateur (A1, en rouge), et (iii) pas de catalyse active mais 3 molécules de H<sub>2</sub>O spectateur (A0, en vert). Les lignes en pointillés correspondent aux calculs C-PCM/UAKS.

## Systèmes tétra-hydratés (PW4)

Les complexes les plus stables, parmi les nombreuses structures explorées sur ce système, et les états de transition correspondants sont présentés sur la Figure V.2.26. Ici encore, il est clair que des complexes tels que **4B1**, **4B2** et **4B4** possèdent des équivalents dans les générations précédentes.

Nous n'avons pas réussi à identifier d'état de transition associé à une catalyse active à deux molécules d'eau (A2) pour l'étape (b) ni à une catalyse totalement passive (A0) pour l'étape (a). Les enthalpies des états de transition que nous avons mis en évidence se situent nettement en dessous de l'enthalpie des réactifs isolés (entre -145 et -206 kJ mol<sup>-1</sup>). Le plus haut en enthalpie (**4a1**, -145 kJ mol<sup>-1</sup>) correspond à la structure de transition de l'étape (a) pour une catalyse active à une molécule d'eau, A1 tandis que le plus bas (**4b1**, -206 kJ mol<sup>-1</sup>) correspond à l'étape (b) du même processus (Figure V.2.27). Comme observé précédemment pour les systèmes de plus petites tailles, la seconde étape est la plus facile pour **PW4-A1**. On notera que la différence d'enthalpie entre les deux états de transition (a) et (b) est du même

ordre de grandeur que celle qui a été calculée pour le système **PW3-A1** (61 contre 75 kJ mol<sup>-1</sup> pour ce dernier) mais plus élevée que celle des systèmes **PW2-A1** et **PW1-A1** (35 et 41 kJ mol<sup>-1</sup>).

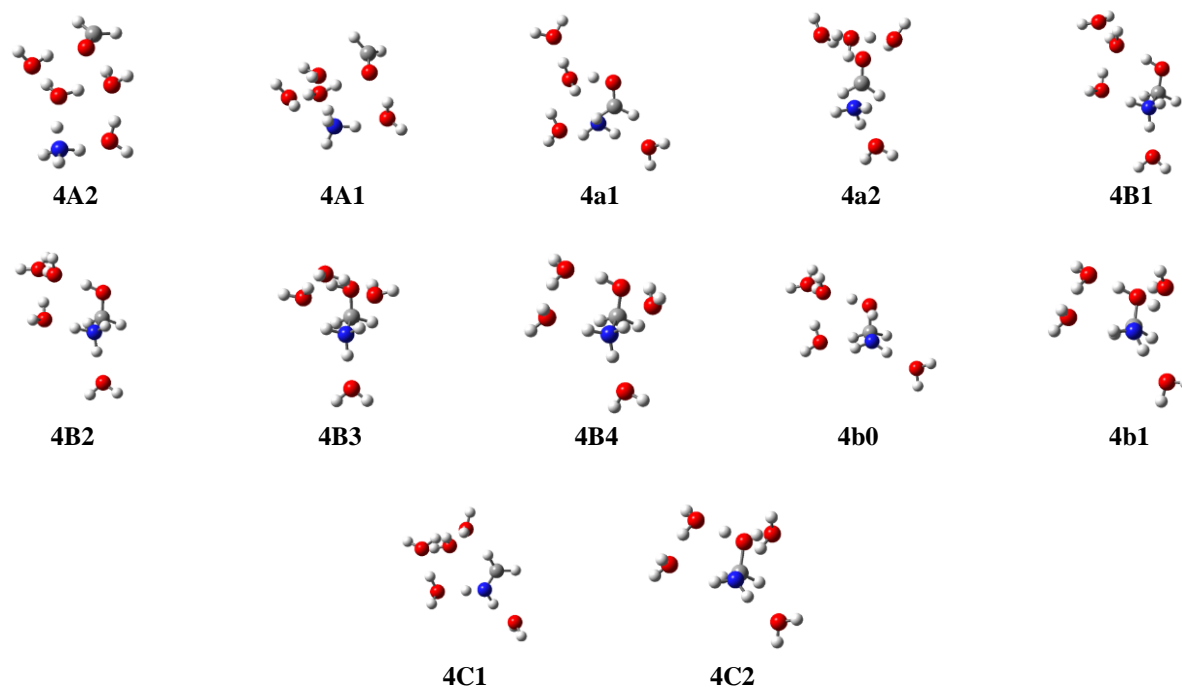


Figure V.2.26. Géométries B3LYP/6-31G(d) des structures **PW4** en phase gazeuse.



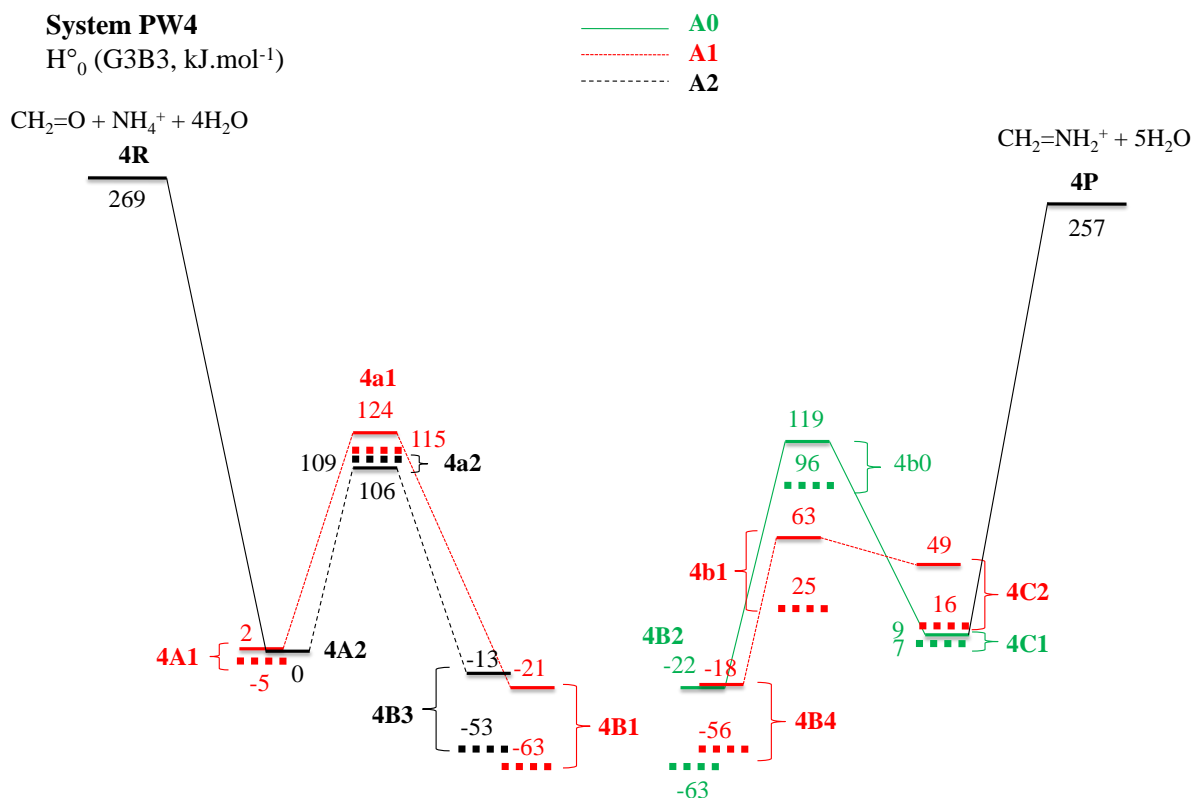


Figure V.2.27. Profil en enthalpie à 0 K en G3B3 pour le système **PW4**. (i) « catalyse active » par 2 molécules de H<sub>2</sub>O plus 2 molécule de H<sub>2</sub>O spectateur (A2, en noir) ; (ii) « catalyse active » par 1 molécule de H<sub>2</sub>O plus 3 molécule de H<sub>2</sub>O spectateur (A1, en rouge), et (iii) pas de catalyse active mais 4 molécules de H<sub>2</sub>O spectateur (A0, en vert). Les lignes en pointillés correspondent aux calculs C-PCM/UAKS.

## Bilan

Nous avons reporté dans la Figure V.2.28 l'ensemble des enthalpies critiques à 0 K associées aux états de transition des étapes (a) et (b) des systèmes **PW0-PW4**, en fonction du nombre total de molécule d'eau. Les principales observations sont les suivantes : Contrairement à ce qui a été constaté avec le milieu neutre, les enthalpies critiques des deux étapes ne diffèrent pas de façon spectaculaire, une sorte de tassement est observée. En tout état de cause, l'étape (a) est toujours l'étape énergétiquement limitante. Le second point à noter est que la catalyse exclusivement passive, de type **A0**, est défavorable par rapport aux catalyses actives **A1** ou **A2**. Ce résultat est comparable à ce qui avait été noté pour les systèmes neutres. Troisième observation essentiellement basée sur les résultats de l'étape (a), la catalyse de type **A2** est toujours plus favorable que la catalyse **A1**. Des valeurs d'enthalpies critiques similaires pour les deux processus avaient été soulignées dans le cas des systèmes neutres, la protonation introduit donc une certaine discrimination. Enfin, dernière remarque, commune aux systèmes

neutres, l'ajout de molécules d'eau spectatrices ne se traduit pas par une réduction évidente des enthalpies critiques.

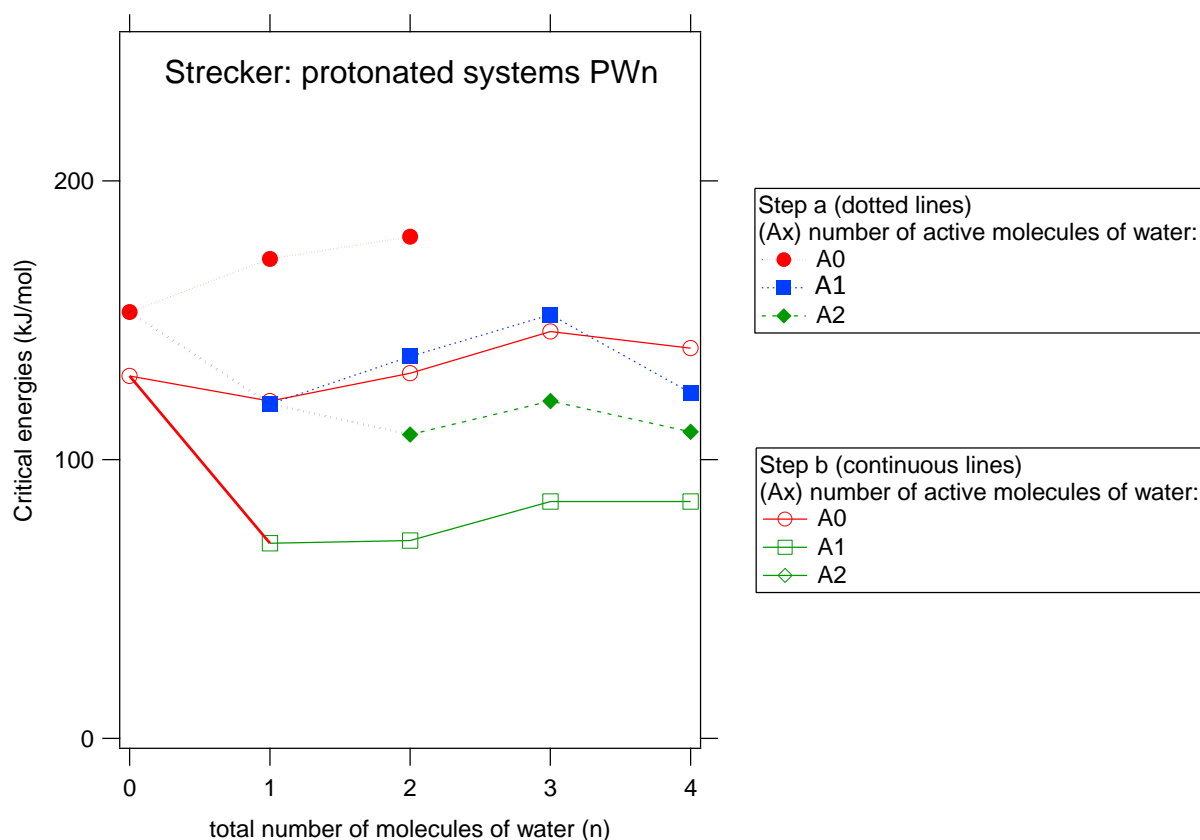


Figure V.2.28. Énergies critiques calculées au niveau G3B3 pour les systèmes protonés PWn comme une fonction du nombre discret de molécule de H<sub>2</sub>O.

La prise en compte des effets de l'environnement au niveau G3B3/C-CPM/UAKS ne modifie pas fondamentalement les conclusions précédentes. En effet, les énergies critiques des barrières des étapes (a) et (b) restent similaires sans ou avec la prise en considération du solvant (voir les Figures V.2.18 à V.2.27).

Les barrières globales associées à la réaction de Strecker en milieu acide (i.e. les différences d'enthalpies entre l'état de transition le plus haut en énergie conduisant à l'imine et les réactifs dans leurs structures complexées les plus stables) sont égales à 153 kJ mol<sup>-1</sup> pour le système **PW0** et est situées entre 105 et 120 kJ mol<sup>-1</sup> pour les systèmes **PW1** à **PW4**. Dans tous les cas, c'est l'étape (a) qui impose l'état de transition le plus haut en énergie. La même tendance est observée lorsque l'on considère le continuum du solvant (eau ou glace). On doit donc en conclure que la réaction de Strecker se heurte, en milieu acide, à une barrière d'une centaine de kJ mol<sup>-1</sup> et que sa réalisation nécessite une activation préalable.

Une source d'activation extérieure est donc nécessaire pour rendre possible la formation du méthylénimine protoné. Il peut être souligné à ce sujet que l'étape de protonation initiale de l'un des réactifs pourrait apporter une quantité d'énergie interne suffisante pour franchir les barrières d'activation, s'il n'y a pas de processus de désactivation résultant de collisions ou d'émission de photons. Dans le milieu interstellaire, l'espèce protonante la plus abondante est l'ion  $\text{H}_3^+$ . L'affinité protonique de  $\text{H}_2$  est égale à  $424 \text{ kJ mol}^{-1}$ . Ainsi, la protonation par  $\text{H}_3^+$  de petites molécules organiques telles que  $\text{H}_2\text{O}$ ,  $\text{H}_2\text{CO}$ ,  $\text{NH}_3$  et  $\text{H}_2\text{CNH}$  (avec des affinité protonique égales à 691, 713, 854 et  $869 \text{ kJ mol}^{-1}$  respectivement) conduit à des formes protonées possédant des énergies internes égales à 267, 289, 430 et  $445 \text{ kJ mol}^{-1}$  respectivement. Par conséquent, sans processus de désactivation, la protonation de  $\text{NH}_3$  par  $\text{H}_3^+$  apporte une énergie interne de  $430 \text{ kJ mol}^{-1}$ , ce qui est suffisamment élevée pour permettre le passage des barrières des étapes (a) et (b) des processus **PW0** à **PW4**. La protonation de l'intermédiaire  $\text{H}_2\text{NCH}_2\text{OH}$  par  $\text{H}_3^+$  est également envisageable. L'affinité protonique calculée pour  $\text{H}_2\text{NCH}_2\text{OH}$  est égale à  $883 \text{ kJ mol}^{-1}$  (G3B3, 0 K), la protonation de cette molécule par l'ion  $\text{H}_3^+$  produirait un ion  $\text{H}_3\text{NCH}_2\text{OH}^+$  porteur d'une quantité d'énergie interne de l'ordre de  $460 \text{ kJ mol}^{-1}$  largement suffisante pour permettre le passage de la barrière (b) ( $\sim 80 \text{ kJ mol}^{-1}$  dans le système **PW1** par exemple) et permettant ainsi de former le méthylénimine protoné.

## V.2.5. Conclusion

De l'ensemble des résultats présentés pour les systèmes neutre **NWn** et protonés **PWn**, la réalisation spontanée de la réaction de Strecker paraît difficile. Un scénario possible dans des échelles de temps très long de la chimie prébiotique pourrait combiner une étape (a) dans un milieu neutre puis une étape (b) dans un environnement protoné. L'ensemble ne nécessiterait qu'une activation modérée, de l'ordre de quelques dizaines de  $\text{kJ mol}^{-1}$ . L'activation par suite d'une protonation par  $\text{H}_3^+$  est également une hypothèse raisonnable même si une partie de l'excédent d'énergie interne apporté par la protonation est dissipé par chocs ou émission de radiations électromagnétiques.

## V.2.6. Bibliographie

- [1] Maruoka, K.; Ooi, T. *Chem. Rev.* **2003**, *103*, 3013-3028.
- [2] Nájera, C.; Sansano, J. M. *Chem. Rev.* **2007**, *107*, 4584-4671.
- [3] Strecker, A. *Ann. Chem. Pharm.* **1850**, *75*, 27-45.
- [4] Moutou, G.; Taillades, J.; Bénéfice-Malouet, S.; Commeyras, A.; Messina, G. ; Mansani, R. *J. Phys. Org. Chem.* **1995**, *8*, 721-730.
- [5] Kobayashi, S.; Ishitani, H. *Chem. Rev.* **1999**, *99*, 1069-1094.
- [6] Gröger, H. *Chem. Rev.* **2003**, *103*, 2795-2827.
- [7] Wang, J.; Liu, X.; Feng, X. *Chem. Rev.* **2011**, *111*, 6947-6983.
- [8] Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. *Chem. Rev.* **2014**, *114*, 285-366.
- [9] Miller, S. L. *Science* **1953**, *117*, 528-529.
- [10] Muñoz Caro, G. M.; Meierhenrich, U. J.; Schutte, W. A.; Barbier, B.; Arcones, Segovia, A.; Rosenbauer, H.; Thiemann, W. H. P.; Brack, A.; Greenberg, J. M. *Nature* **2002**, *416*, 403-406.
- [11] Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. *Nature* **2002**, *416*, 401-403.
- [12] Lerner, N. R.; Cooper, G. W. *Geochim.Cosmochim.Acta* **2005**, *69*, 2901-2906.
- [13] Brack, A. *Chem. & Biodiversity* **2007**, *4*, 665-679.
- [14] Elsila, J. E.; Dworkin, J. P.; Bernstein, M. P.; Martin, M. P.; Sandford, S. A. *ApJ* **2007**, *660*, 911-918.
- [15] Nuevo, M.; Auger, G.; Blanot, D.; d'Hendecourt, L. *Orig. Life Evol. Biosph.* **2008**, *38*, 37-56.
- [16] Plankensteiner, K.; Reiner, H.; Schranz, B.; Rode, B. M. *Angew. Chem. Int. Ed.* **2004**, *43*, 1886-1888.
- [17] Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L. Bada, J. L. *Orig. Life Evol. Biosph.* **2008**, *38*, 105-115.
- [18] Parker, E.; Cleaves, H. J.; Callahan, M. P.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Orig. Life Evol. Biosph.* **2011**, *41*, 201-212.
- [19] Aubrey, A. D.; Cleaves, H. J.; Bada, J. L. *Orig. Life Evol. Biosph.* **2009**, *39*, 91-108.
- [20] Chyba, C.; Sagan, C. *Nature*, **1992**, *355*, 125-132.
- [21] Pizzarello, S.; Huang, Y.; Becker, L.; Poreda, R. J.; Nieman, R. A.; Cooper, G.; Williams, M. *Science* **2001**, *293*, 2236-2239.
- [22] Pizzarello, S. *Acc. Chem. Res.* **2006**, *39*, 231-237.
- [23] Pizzarello, S.; Shock, E. *Cold Spring Harbor Perspect.Biol.* **2010**, *2*, a002105 1-19.
- [24] Burton, A. S.; Stern, J. C.; Elsila, J. E.; Glavin, D. P.; Dworkin, J. P. *Chem. Soc. Rev.* **2012**, *41*, 5459-5472.
- [25] Degens, E. T.; Bajor, M. *Naturwissenschaften* **1962**, *49*, 605-606.
- [26] Elsila, J. E.; Glavin, D. P.; Dworkin, J. P. *Meteorit.Planet. Sci.* **2009**, *44*, 1323-1330.
- [27] Thaddeus, P. *Phil. Trans. R. Soc. B* **2006**, *361*, 1681-1687.
- [28] Geppert, W. D.; Larsson, M. *Chem. Rev.* **2013**, *113*, 8872-8905.
- [29] Woon, D. E. *Icarus*, **1999**, *142*, 550-556.
- [30] Walch, S. P.; Bauschlicher Jr., C. W.; Ricca, A.; Bakes, E. L. O. *Chem. Phys. Lett.* **2001**, *333*, 6-11.
- [31] Minyaev, R. M.; Lepin, E. A. *Mendeleev Commun.* **1997**, *7*, 189-191.
- [32] Courmier, D.; Gardebien, F.; Minot, C.; St-Amant, A. *Chem. Phys. Lett.* **2005**, *405*, 357-363.
- [33] Feldmann, M. T.; Widicus, S. L.; Blake, G. A.; Kent, D. R.; Goddard, W. A. *J. Chem. Phys.* **2005**, *123*, 034304 1-6.
- [34] Rimola, A.; Sodupe, M.; Ugliengo, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5285-5294.

- [35] Johari, G. P.; Whalley, E. *J. Chem. Phys.* **1981**, 75, 1333-1340.
- [36] Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, 1, 70-77.
- [37] Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, 109, 10795-10797.
- [38] Bouchoux, G.; Hoppilliard, Y. *Int. J. Mass Spectrom. Ion Processes* **1987**, 75, 1-14.
- [39] Heerma, W.; Kulik, W.; Burgers, P. C.; Terlouw, J. K. *Int. J. Mass Spectrom. Ion Processes* **1988**, 84, R1-R5.
- [40] Uggerud, E. *J. Chem. Soc., Perkin Trans. 2* **1996**, 1915-1920.
- [41] Okada, S.; Abe, Y.; Taniguchi, S.; Yamabe, S. *J. Am. Chem. Soc.* **1987**, 109, 295-300.

## Conclusions générales et perspectives

L'axe conducteur de cette thèse était le rôle des interactions non-covalentes sur des propriétés intrinsèques (structurale, thermochimique et électronique) et la réactivité de peptides. Pour en avoir une meilleure compréhension, nous nous sommes intéressés à des peptides isolés et des clusters de tailles croissantes en utilisant les outils de la chimie quantique et de la modélisation moléculaire. Au bilan, toutes nos études, réalisés sur des radicaux cations isolés, sur des molécules chargées micro-hydratés, et sur des clusters protonés, confirment que les interactions non-covalentes jouent un rôle clé dans la chimie de ces systèmes.

Dans plusieurs cas, une étape préliminaire est nécessaire avant de pouvoir aborder l'étude de ces interactions : l'exploration de la surface d'énergie potentielle pour déterminer la (ou les) structure(s) des systèmes à analyser. Nos travaux ont illustré la difficulté de cette recherche de structure et la nécessité de l'adapter au système étudié. Dans le cas de la micro-hydratation des acides aminés, nous avons mis au point une stratégie de modélisation combinant (i) une exploration de la surface d'énergie potentielle par AMOEBA, (ii) de nouveaux calculs d'énergies relatives sur tous les conformères AMOEBA et/ou d'optimisation de géométries avec des méthodes semi-empirique (PM6 et (ou) PM6-DH2) et (ou) DFT (M06 ou B3LYP), (iii) la construction d'un arbre de Darwin et (iv) l'identification des structures manquantes par l'analyse des descendance. Cette stratégie s'est révélée particulièrement efficace pour obtenir la population des conformères dans une gamme d'énergie choisie. Cette stratégie a dû être modifiée dans le cadre de l'étude des clusters de glycine bêtaïne protonés car nous avons choisi de travailler avec des clusters représentatifs de chaque famille identifiée au cours de la dynamique moléculaire et non pas avec une population de conformères.

Dans le premier chapitre, l'étude des  $\text{AcAAH}^+(\text{H}_2\text{O})_n$  nous a permis de montrer que le squelette peptidique n'est pas nécessairement préservé par l'ajout de seulement quelques molécules de  $\text{H}_2\text{O}$ , illustrant l'influence de la solvation sur la structure tridimensionnelle des peptides. L'ajout contrôlé d'une molécule d'eau sur différents sites nous a permis de classer les sites préférentiels d'attaches pour les différentes générations  $\mathbf{G}_n$ . De plus, les groupements localisés aux environs du site d'hydratation influencent les interactions  $\text{H}_2\text{O} \cdots \text{site}$  et réciproquement. Nous avons également pu reproduire les énergies d'hydratation expérimentales. La confrontation théorie-expérience nous a permis de conclure que les

données thermochimiques expérimentales seules sont utiles mais pas suffisantes pour identifier les sites de localisation des premières molécules de H<sub>2</sub>O.

En perspective, il serait intéressant d'étudier l'effet des molécules d'eau sur les propriétés structurales et thermochimiques des acides aminés neutres N $\alpha$ -acétylés notés [NAX]. Puisque [NAH + H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub>, [NAR + H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub> et [NAK + H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub> ont déjà été étudiés dans nos travaux, les systèmes neutres choisis seraient [NAH](H<sub>2</sub>O)<sub>n</sub>, [NAR](H<sub>2</sub>O)<sub>n</sub> et [NAK](H<sub>2</sub>O)<sub>n</sub>. Il ne sera pas possible d'avoir des données expérimentales en spectrométrie de masse, mais cette étude permettrait d'analyser l'évolution de la nature des sites préférentiels d'hydratation (cycle imidazole, NH<sub>3</sub>, guanidine, COOH par exemple) par rapport au système [NAX + H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub>. De plus, l'ajout d'un proton sur le système [NAX](H<sub>2</sub>O)<sub>n</sub> conduit-il aux structures déjà identifiées de [NAX + H]<sup>+</sup>(H<sub>2</sub>O)<sub>n</sub> dans ce travail de thèse, ou est-ce que de nouvelles structures avec un (des) nouveau(x) site(s) de protonation (comme le groupement acyle) seront identifiées ?

Avec l'étude des systèmes [M + 2H]<sup>2+</sup>, nous voulions comprendre l'effet des liaisons hydrogène sur la structure électronique des intermédiaires radicaux cations formés avant les processus de fragmentation en ECD et (ou) ETD. Après avoir testé l'efficacité de plusieurs fonctionnelles, conventionnelles et RSHs, basée sur des critères énergétiques et des densités de spin, nous avons pu montrer que LC-BLYP et  $\omega$ B97 sont des fonctionnelles adaptées pour établir les structures électroniques de ces intermédiaires. Elles possèdent 100 % d'échange HF à longue distance inter-électronique et les valeurs de leur paramètre d'atténuation sont suffisamment élevées pour éviter les problèmes de distribution des densités de spin et une sous-estimation de leurs énergies d'excitations verticales. Ensuite, nous avons pu montrer que le nombre de liaison hydrogène est un facteur déterminant de la structure électronique de ces intermédiaires. En effet, les liaisons hydrogène autour des ammoniums influent sur le nombre de  $\sigma_{N-H}^*$  disponible, la valeur des ERs et la relaxation géométrique.

Tout comme les sites d'hydratation des AcAAH<sup>+</sup>, le site de capture électronique de l'électron est également influencé par la présence de liaisons hydrogène autour des sites chargés. Ainsi, les sites d'hydratation et de capture électronique pourraient servir de sonde indirecte de la structure 3D des peptides.

Dans un peptide radical cation contenant plusieurs ammoniums, l'électron se localise préférentiellement dans les  $\sigma_{N-H}^*$  et  $\pi_{CO}^*$  associées à l'ammonium le moins solvato par des carbonyles à condition que les ERs des ammoniums ne soient pas proches. Les sites n'ayant pas capturé l'électron sont peu influencés par la présence de l'électron célibataire au cours de

la relaxation géométrique. Lors du réarrangement de l'intermédiaire radical cation, les processus d'élimination d'un H<sup>•</sup> et de transfert couplé proton-électron possèdent de faibles barrières ( $\Delta G^\circ_{298} \sim 0.1$  eV et  $< 0.5$  eV, respectivement). Toutefois, l'élimination d'un radical H est favorisée d'un point de vue cinétique et thermodynamique par rapport au transfert couplé proton-électron justifiant ainsi qu'il soit le processus le plus observé expérimentalement. L'intermédiaire radical cation n'est pas nécessairement un point stationnaire sur la surface d'énergie potentielle, le transfert couplé proton-électron pouvant se faire spontanément si la délocalisation de la densité de spin de  $\sigma_{N-H}^*$  vers la  $\pi_{CO}^*$  est déjà bien initié. L'étude de la relaxation des états excités peut être envisagée dans de futurs travaux.

L'étude des clusters de glycine-bétaïne avait pour objectif d'identifier leurs structures et de comprendre certains de leur processus de dissociation. Par exemple, l'élimination d'un neutre et la séparation de charges ont été observées expérimentalement en CID et en EID pour des électrons d'énergie supérieure à 12 eV. Après avoir généré les structures, nous avons pu mettre en évidence une compétition entre les interactions électrostatiques attractives et répulsives. La prédominance de l'un de ces deux types d'interaction dépend fortement de la taille des clusters et permet d'expliquer leur existence ou non en phase gazeuse. De plus, il existe une multitude de liaison hydrogène de type  $CO \cdots HC$ , en plus des deux liaisons  $COOH \cdots OCO$ , permettant une meilleure cohésion pour les clusters de grandes tailles. Ainsi, que ce soit pour les  $AcAAH^+(H_2O)_n$  et les clusters de glycine-bétaïne, les liaisons hydrogène sont un élément essentiel à leur structuration.

La voie de fragmentation associée à l'élimination d'une GB neutre est énergétiquement similaire pour les différentes tailles de clusters à partir de  $n > 4$  puisque les mêmes types d'interactions sont rompus. L'explosion coulombienne et l'élimination d'une GB neutre sont deux processus compétitifs pour  $[GB_{13} + 2H]^{2+}$ . Cela permet d'expliquer que ces deux processus soient observés expérimentalement en EID à partir des mêmes énergies d'électrons.

Les premières étapes de la synthèse de Strecker sont facilitées par un environnement local de  $H_2O$ , comme cela a déjà été mentionné de nombreuses fois dans la littérature. Comme pour  $AcAAH^+$ , les molécules de  $H_2O$  sont principalement en interactions avec le site chargé. Le rôle de l'eau dans la catalyse est important. Néanmoins, l'augmentation du nombre de molécules actives ne mène pas nécessairement à la diminution des barrières d'activation. Généralement, une molécule de  $H_2O$  active est suffisante.



En perspective, il serait intéressant d'étudier les propriétés électroniques des clusters de glycine bétaine multichargés afin de comprendre les processus de fragmentation tels que la rupture des liaisons covalentes et l'élimination simultanée de mGB et  $H^\bullet$ . À la suite de nos travaux menés au chapitre IV, les fonctionnelles LC-BLYP et  $\omega$ B97 pourront être utilisées dans cette future étude. L'effet de la capture électronique sur le complexe de départ de la réaction de Strecker  $NH_3 (NH_4^+) + H_2CO + nH_2O$  peut également être envisagé afin d'étudier l'évolution des barrières critiques sur les transferts d'hydrogène ( $H^+$  et (ou)  $H^\bullet$ ) au cours d'une synthèse type Strecker dans des conditions prébiotiques.

# Production scientifique

## Articles Scientifiques :

- 1) V. Riffet, G. Frison, G. Bouchoux,  
*Acid-base thermochemistry of gaseous oxygen and sulfur substituted amino acids (Ser, Thr, Cys, Met)*, Phys. Chem. Chem. Phys., **13**, 18561-18580 (2011).
- 2) V. Riffet, S. Bourcier, G. Bouchoux,  
*Gas-phase basicity and acidity of tryptophan*,  
Int. J. Mass Spectrom., **316-318**, 47-56 (2012).
- 3) V. Riffet, G. Bouchoux,  
*Gas-phase structures and thermochemistry of neutral histidine and its conjugated acid and base*, Phys. Chem. Chem. Phys., **15**, 6097-6106 (2013).
- 4) V. Riffet, D. Jacquemin, E. Cauët, G. Frison,  
*Benchmarking DFT and TD-DFT functionals for the ground and excited states of hydrogen-rich peptide radicals*,  
J. Chem. Theory Comput. **DOI**: 10.1021/ct5004912
- 5) V. Riffet, G. Bouchoux, G. Frison,  
*Microhydration of protonated N $\alpha$ -acetylhistidine: a theoretical approach*,  
en préparation
- 6) V. Riffet, G. Bouchoux, G. Frison,  
*Structures of hydrated protonated N $\alpha$ -acetyllysine: exploration of the conformational landscape*,  
en préparation
- 7) V. Riffet, G. Bouchoux, G. Frison,  
*Reciprocal influences of non-covalent interactions of microhydrated protonated acetyl amino acids*,  
en préparation
- 8) V. Riffet, G. Frison, R. Pascal, G. Bouchoux,  
*Quantum-chemical modeling of the first step of the Strecker synthesis of glycine: from the gas-phase to water solvation and catalysis*,  
en préparation

## Communications :

### Orales :

- 1) *Structure électronique des peptides radicaux cations : performances de la DFT et effets des liaisons hydrogène*, Journées « Modélisation à Paris », les 18 et 19 septembre 2014, Paris, France.  
Vanessa Riffet, Denis Jacquemin, Emilie Cauët, Gilles Frison

2) *Les liaisons hydrogène, un facteur clé de la structure électronique des peptides radicaux cations*, GDR « Edifices Moléculaires Isolés et Environnés », du 06 au 09 octobre 2013, Ecully, France.

Vanessa Riffet, Emilie Cauët, Denis Jacquemin, Gilles Frison

3) *Exploration de la flexibilité conformationnelle d'acides aminés protonés microhydratés*, Journées « Modélisation à Paris », les 14 et 15 juin 2012, Paris, France.

Vanessa Riffet, Guy Bouchoux, Gilles Frison.

## **Posters :**

1) *A computational approach to microhydration of protonated amino acids and small peptides*, Conference on Isolated Biomolecules and Biomolecular Interactions, du 22 au 27 janvier 2012, Les Diablerets, Suisse.

Vanessa Riffet, Guy Bouchoux, Denis Jacquemin, Gilles Frison.

2) *Influence réciproques des interactions non-covalentes sur des acides-aminés protonés microhydratés*, XIII<sup>ème</sup> Rencontre des Chimistes Théoriciens Francophones, du 1<sup>er</sup> au 5 juillet 2012, Marseille, France.

Vanessa Riffet, Guy Bouchoux, Gilles Frison.

3) *Environmental effects on the electronic structure of peptide radical cations*, TD-DFT Conference, du 23 au 26 avril 2013, Nantes, France.

Vanessa Riffet, Emilie Cauët, Denis Jacquemin, Gilles Frison.

4) *Clusters de glycine bêtaïne: structures et énergies d'interactions*, GDR « Edifices Moléculaires Isolés et Environnés », du 06 au 09 octobre 2013, Ecully, France.

Vanessa Riffet, Gilles Frison, Emilie Cauët, Richard O'Hair, Linda Feketeova.

5) *Environmental effects on the electronic structure of hydrogen-rich peptide-cation radicals*, Conference on Isolated Biomolecules and Biomolecular Interactions, du 18 au 23 mai 2014, Porquerolles, France.

Vanessa Riffet, Emilie Cauët, Denis Jacquemin, Gilles Frison.

6) *Structure électronique et environnement local des peptides radicaux cations*, XIV<sup>ème</sup> Rencontre des Chimistes Théoriciens Francophones, du 30 au 4 juillet 2014, Paris, France.

Vanessa Riffet, Emilie Cauët, Denis Jacquemin, Gilles Frison.

La structure tridimensionnelle et les propriétés physico-chimiques des biomolécules telles que les peptides sont gouvernées non seulement par leur composition primaire mais aussi par diverses interactions non-covalentes, inter- et intra-moléculaires. La description, la mesure et les effets de ces interactions sont actuellement au cœur de nombreux travaux de recherches à l'interface entre la biologie et la physico-chimie. Dans ce cadre, nos travaux de thèse visent à une meilleure compréhension de ces interactions au sein de biomolécules et d'agrégats en utilisant les outils de la chimie quantique et de la modélisation moléculaire. Pour cela, du fait de la complexité et de la taille des systèmes biologiques réels, des modèles chimiques simplifiés ont été élaborés. Ceux-ci nous ont permis d'étudier et de comprendre la nature et les effets de ces interactions prises « individuellement » sur la structure géométrique et électronique de systèmes moléculaires tels que de petits peptides et des agrégats de la glycine bêtaïne. Le rôle des interactions non-covalentes sur la réactivité dans des conditions physico-chimiques variées (réduction électronique, phase gazeuse, microsolvation) constitue également une des approches de notre travail.

Parmi les systèmes étudiés, nombres d'entre eux comportent un ou plusieurs groupes chargés ammonium, guanidinium ou imidazolium, présents dans les chaînes latérales des acides aminés lysine, arginine et histidine. Nous avons montré que l'environnement chimique et électronique de ces groupes chargés influence largement la structure et la réactivité des molécules qui les contiennent. Enfin, au cours de nos travaux, nous avons réalisé une calibration des méthodes de modélisation pour l'étude des propriétés électroniques de peptides radicalaires et mis au point des protocoles d'exploration de surface d'énergie potentielle à partir du champ de force polarisable AMOEBA, ceci afin d'optimiser l'exhaustivité des recherches conformationnelles pour des systèmes biologiques flexibles.

The three-dimensional structure and physico-chemical properties of biomolecules such as peptides are not only governed by their elementary composition but also various non-covalent intra-and inter-molecular interactions. The characterization, measurement and effects of these interactions are currently at the center of many researches at the interface between biology and physical chemistry. In this context, the aim of our thesis is a better understanding of these interactions in biomolecules and aggregates using the tools of quantum chemistry and molecular modeling. In this regard, due to the complexity and size of the real biological systems, chemical models have been developed. These have allowed us to study and understand the nature and effects of these interactions taken "individually" on the geometric and electronic structure of molecular systems such as small peptides and aggregates of glycine betaine. The role of non-covalent interactions on the reactivity of various physico-chemical conditions (electronic reduction, gas phase, microsolvation) is also one of the several approaches of our work.

Among the systems examined, many of them included at least one ammonium, guanidinium or imidazolium charged groups, which are founded in the side chains of the amino acids lysine, arginine and histidine. We have shown that the chemical environment of these electronic charges greatly influences the structure and reactivity of the molecules that contain these groups. Eventually, we both performed a calibration of modeling methods for the study of the electronic properties of radical peptides and developed new protocols for the potential energy surface exploration starting from the AMOEBA polarizable force field, in order to optimize the conformational searches exhaustivity for flexible biological systems.